MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Who wears the trousers?

A cognitive linguistic approach to gender bias in idioms"

verfasst von / submitted by

Fiona Rosilda Tamminga

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

Master of Arts (MA)

Wien, 2023 / Vienna, 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 812
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	MA English Language and Linguistics
Betreut von / Supervisor:	Univ.-Prof. Dr. Mathilde Eveline Keizer

Abstract (English)

Research has shown that proverbs and idioms can be considered as a source of sexism. However, they do not always portray this in an obvious manner, rather using neutral language as a disguise. As proverbs and idioms are considered to be influential in building and controlling stereotypes, it is apparent that awareness of these ideologies is important. The main purpose of this thesis is to investigate the connection between these innocent-looking idioms and their impact on gender bias. This study aims to answer the following research questions: "Do outwardly neutral-looking idioms contain a hidden gender bias?", "Are native speakers more aware of this hidden bias compared to non-native speakers?", "Is the gender bias more pronounced in the results from idioms containing hostile sexism compared to those containing benevolent sexism?", and "Are the results from the participants that identify as female significantly different from those that identify as male?". The classification of the idioms is done according to the Ambivalent Sexism Theory by Glick and Fiske (1996, 2010). The hypothesis is based on the reasoning that the key content word in the idiom will have been used and / or is being used in either a more masculine or feminine context, thereby infusing the idiom with a hidden bias. Diachronic corpus research and an online questionnaire show that there is a general male bias in the idioms of this study; however, the dataset was not sufficient to support a general claim in this area. Furthermore, the groups that were tested, namely a female, male, native, and non-native group, did not portray significant differences between their answers. Finally, the results do support the Ambivalent Sexism Theory by showing a significant difference between the hostile and benevolent idioms. To conclude, even though this experiment did not produce the expected results, the study did highlight the importance of research in this area; therefore, it is recommended to continue investigating gender bias in idioms.

Abstrakt (Deutsch)

Untersuchungen haben gezeigt, dass Sprichwörter und Redewendungen als Quelle von Sexismus angesehen werden können. Sie stellen dies jedoch nicht immer auf eine offensichtliche Weise dar, sondern verwenden eher eine neutrale Sprache als Tarnung. Da Sprichwörter und Redewendungen als einflussreich für den Aufbau und die Kontrolle von Stereotypen angesehen werden, ist es offensichtlich, dass das Bewusstsein für diese Ideologien wichtig ist. Das Hauptziel dieser Arbeit ist es, die Verbindung zwischen diesen unschuldig aussehenden Redewendungen und ihren Einfluss auf die geschlechtsspezifische Voreingenommenheit zu untersuchen. Diese Studie zielt darauf ab, die folgenden Forschungsfragen zu beantworten: „Enthalten nach außen hin neutral wirkende Redewendungen eine versteckte geschlechtsspezifische Voreingenommenheit?“, „Sind Muttersprachler dieser versteckten Voreingenommenheit stärker bewusst als Nicht-Muttersprachler?“, „Ist die geschlechtsspezifische Voreingenommenheit stärker ausgeprägt in den Ergebnissen von Redewendungen, die feindseligen Sexismus enthalten, im Vergleich zu denen, die wohlwollenden Sexismus enthalten?“ und „Unterscheiden sich die Ergebnisse der Teilnehmer, die sich als weiblich identifizieren, signifikant von denen, die sich als männlich identifizieren?“. Die Einordnung der Redewendungen erfolgt nach der Ambivalent Sexism Theory von Glick und Fiske (1996, 2010). Die Hypothese basiert auf der Überlegung, dass das Schlüsselinhaltswort in der Redewendung entweder in einem eher männlichen oder weiblichen Kontext verwendet wurde und / oder verwendet wird, wodurch die Redewendung mit einer versteckten Voreingenommenheit versehen wird. Diachrone Korpus Recherchen und ein Online-Fragebogen zeigen, dass es eine allgemeine männliche Tendenz in den Idiomen dieser Studie gibt; der Datensatz reichte jedoch nicht aus, um eine allgemeine Behauptung in diesem Bereich zu stützen. Darüber hinaus zeigten die getesteten Gruppen, nämlich eine weibliche, männliche, muttersprachliche und nicht-muttersprachliche Gruppe, keine signifikanten Unterschiede zwischen ihren Antworten. Schließlich stützen die Ergebnisse die Ambivalent Sexism Theory, indem sie einen signifikanten Unterschied zwischen den feindseligen und wohlwollenden Redewendungen zeigen. Abschließend lässt sich sagen, dass dieses Experiment zwar nicht die erwarteten Ergebnisse erbrachte, die Studie jedoch die Bedeutung der Forschung auf diesem Gebiet hervorhob; daher wird empfohlen, die geschlechtsspezifische Voreingenommenheit in Redewendungen weiter zu untersuchen.

Tables and figures

Tables

Figures

Table of contents

Introduction

Research demonstrates that proverbs are a source of sexism in many languages and speech communities. An example of such a proverb can be found in the title of Schipper's (2010) research "Never marry a woman with big feet: women in proverbs around the world". Additionally, "[g]irls need men or walls", or "[a] man of straw is worth as much as a woman of gold" display a clear gender bias (Lomotey and Chachu 2020:73-74). However, proverbs do not always exhibit their gender bias that openly. Lomotey (2019:161) recognizes that misogynistic ideas that were propagated through proverbs with overt sexism in Peninsular Spanish are now distributed through "refashioned" forms that are not as obviously sexist.

It is therefore necessary to investigate the influence of these oftentimes innocent-looking phrases. Lomotey and Chachu (2020:70) argue that "[t]he insistent, tenacious and insidious nature of gender ideologies and stereotypes have contributed to maintaining the prevailing gender status quo and have consequently perpetuated gender discrimination in many cultures and across several generations". Furthermore, the idioms of a certain culture will portray their "underlying ideas, principles and values" (Nezhelskaya, Samarina and Shibkova 2018:1). Essentially, "language reflects, records and transmits social, cultural and gender differences" (Nezhelskaya, Samarina and Shibkova 2018:1). Indeed, Lomotey and Chachu (2020:71) underline that proverbs are influential mechanisms that build and control ideologies as well as stereotypes. They are seen as "embodiments of collective wisdom" with a stature that is unquestionable (Lomotey and Chachu 2020:71).

Schippers (2004:13) emphasizes that it is important to be aware of sexist language and to not follow China's example by reprinting books on proverbs without sexist content. These proverbs could be used to increase awareness about sexism, because they are essentially a part of the culture and will have influenced the members of its society (Schippers 2004:13). Such negative messages about women impact "policy makers" and thereby hinder any progression. In order to change this situation, it is vital to investigate these items of our language (Schippers 2004:13). Visions and ideologies change over time and increasing awareness on this issue will only be beneficial (Schippers 2004:13).

This thesis focusses on idioms that do not show their true colours immediately: they are inwardly sexist, but have a neutral appearance. Indeed, Lomotey and Chachu (2020:70) claim that misogynistic ideologies are often concealed in outwardly neutral language use (Lomotey and Chachu 2020:70). A clear example of this is the idiom *to wear the trousers* or *to wear the pants*. The meaning of this idiom is of course *to be in charge*. However, the way in which it is implemented in our society has a denigrating element to it in that it usually denotes women that have challenged the societal norm of the men being in power. On the surface, this idiom is not as overtly sexist as the proverb warning men from marrying women with big feet; nevertheless, assuming that the power in our society belongs solely to men and any woman challenging this should be ridiculed does not portray gender equality.

The previously mentioned research has already shown that there is an interest in studies on sexism in proverbs. However, it is also clear that the link between the "outwardly neutral language" that Lomotey and Chachu (2020:70) speak of and their influence on gender bias has not been fully revealed yet. For this MA thesis, I would like to explore whether neutral looking expressions influence gender bias and to what extent. In order to achieve this, I would like to move away from proverbs, which have had some attention already, and focus on idioms that do not show any obvious outward bias. I have found little to no research in this area; thus providing me with the opportunity to start to close this gap in the existing knowledge. Moreover, I will include both native and non-native speakers of English in my research, something that also has not been a main focus of idiom research yet.

This study will aim to answer the following research questions: "Do outwardly neutral-looking idioms contain a hidden gender bias?", "Are native speakers more aware of this hidden bias compared to non-native speakers?", "Is the gender bias more pronounced in the results from idioms containing hostile sexism compared to those containing benevolent sexism?", and "Are the results from the participants that identify as female significantly different from those that identify as male?". The classification of the idioms is done according to the Ambivalent Sexism Theory by Glick and Fiske (1996, 2010).

The hypothesis of this study is based on the reasoning that the key content word in the idiom, such as *trousers* in *who's wearing the trousers,* will have been used and / or is being used in either a more masculine or feminine context, thereby infusing the idiom with a hidden bias. This gender bias can hopefully be made visible through diachronic corpus research with graphs detailing the development throughout the decades 1820-2010 of the use of *his / her [noun]* or *his / her [verb]* if the keyword is a noun or a verb respectively. Additionally, the results from an online questionnaire will provide an insight into the minds of both native and non-native speakers.

The general expectation is that there is indeed a hidden bias in the idioms of this study. *Trousers* / *pants* and the other keywords will probably be employed in either a more masculine or a more feminine context, thereby enforcing the bias. Furthermore, it is also expected that native speakers will show an equal bias compared to non-native speakers. Schippers (2010:14) points out that people have the ability to recognize the meaning of proverbs about women from other cultures (Schippers 2010:14). Therefore, the participants will probably understand the hidden bias, just as proverbs from other cultures, because they portray a universal concept. Moreover, Beck and Weber (2016:12) claim that non-native speakers with a high proficiency in their target language process figurative meanings similarly to native speakers of that language. Additionally, it is anticipated that hostile, or overt, forms of sexism will be more easily detected by the participants of the questionnaire, while benevolent, or covert, forms might not be recognized as being sexist and will thus remain hidden. Finally, it is predicted that the answers from the female group of participants will likely differ from the male group when distinguishing overt and covert sexism. People identifying as women are probably more sensitive to both types due to them being the victims of sexism. Most people identifying as men will no doubt be aware of it; however, as their experiences are indirect, their sensitivity will presumably not be as high.

This thesis will answer the aforementioned questions in six chapters. The first chapter will provide the theoretical background and the relevant concepts. After this, the second chapter will explain the research methodology. Following, the third chapter will give detailed information about the statistical analysis of this thesis. In addition, the fourth chapter will continue with the diachronic analysis based on the Corpus of Historical American English and Excel templates. The penultimate chapter will provide a discussion of these results. Finally, the sixth chapter will contain the conclusion.

1. Theoretical background and relevant concepts

1.1 Idioms

Idioms, or fixed expressions generally, have hardly been the centre of attention in standard research on language production even though they cannot be considered as exceptions in the language (Sprenger, Levelt and Kemper 2006:162). Pollio et al. (1977) estimate that the number of figurative expressions that are produced during a minute of speech averages four (Cieślicka 2006:115). Furthermore, Jackendoff (1995) estimates that the number of fixed expressions "including names, titles, poems, and the like" that people know is at least equal to the number of words that are in their vocabulary (Sprenger, Levelt and Kemper 2020:162). Even though the frequencies that can be found are usually estimated, not actually counted, and the numbers go from as low as tens of thousands to as high as several hundred thousands (Sprenger 2003:5), it is easy to see the relevance of studying idioms for linguistic and psycholinguistic purposes (Cieślicka 2006:115). Indeed, studying idioms can shine a figurative light on the processes that are involved in language comprehension, because they are understood with ease by language users even though their meaning is oftentimes opaque (Cieślicka 2006:116).

1.2 Definition

The terminology in the field of fixed expressions and idioms is "problematic" according to Moon (1998:2); there have been discussions on the topic, but so far there has been no agreement on it. Different terms are used to describe the same phenomena, while the same terms are used to describe completely different things (Moon 1998:2). This can for instance be seen in Schippers (2010:20) who categorizes *wearing the pants* as a proverb instead of an idiom as is done in this thesis. It is thus essential to start with a general introduction into the field.

Fixed expression is a broad term covering several "holistic units of two or more words" such as "frozen collocations, grammatically ill-formed collocations, proverbs, routine formulae, sayings, similes", and idioms (Moon 1998:2). Moon (1998:6) uses three criteria to assess whether a combination of words can be recognized as a fixed expression or not: "institutionalization, lexicogrammatical fixedness, and non-compositionality". First, institutionalization is the process through which the formation of the lexical item, in this case the fixed expression, is acknowledged and accepted as being part of the language (Bauer 1983:48). Second, lexicogrammatical fixedness indicates that the unit is defect in some degree (Moon 1998:7). This can be seen in the restricted use of a different "aspect, mood, or voice" in for instance "*call the shots*" or "*shoot the breeze*" (Moon 1998:7). However, Moon (1998:7) cautions that fixedness is complex; either institutionalization or the repeating occurrence of a fixed unit does not mean that this particular string of words can be called a fixed expression. Conversely, not all fixed expressions are completely frozen (Moon 1998:7). Thus, institutionalization and fixedness need one more criterium: non-compositionality. A combination of words is considered non-compositional when the meaning that is derived from the interpretation of the single words of this string does not equal the institutionalized meaning (Moon 1998:8). Finally, the aforementioned criteria are variable; they are present in all fixed expressions, but can vary in the degree of their presence (Moon 1998:8). Indeed, Moon (1998:23) cautions that the typology of these fixed expressions is not completely satisfactory, because the separate categories overlap and one expression might belong to more than a single category.

Idiom as a term is ambiguous, containing two meanings (Moon 1998:3). Firstly, "idiom is a particular manner of expressing something in language, music, and so on, which characterizes a person or group" (Moon 1998:3). Secondly, "idiom is a particular lexical collection or phrasal lexeme, peculiar to a language" (Moon 1998:3). Of course this turns out to be confusing, because idiom is used here first as a superordinate and second as a subordinate term (Moon 1998:4). Further opposing views can be seen in the narrow and broad use of the term idiom. The narrower use reduces idiom to be a "fixed and semantically opaque or metaphorical" unit, or in other words, "not the sum of its parts" (Moon 1998:4). These are also called "pure idioms" (Moon 1998:4). Examples of these are "*kick the bucket* or *spill the beans*" (Moon 1998:4). On the other hand, the broader use sees idiom as a term that can be generally applied to many types of multi-word combinations, whether they are semantically opaque or not (Moon 1998:4).

In my thesis, I will follow Moon's (1998) second definition, thereby also following for instance Swinney and Cutler (1979:523), who say that "[i]n its simplest form an idiom is a string of two or more words for which meaning is not derived from the meaning of the individual words comprising that string" and Libben and Titione (2008:1103), who define idioms as "phrases whose figurative meanings are distinct from their component words". This means that I will not narrow down the meaning of idiom to that of the pure idiom only. This coincides with Moon's (1998:4) warning that grammatically abnormal units "such as *by and large*", "transparent metaphors such as *skate on thin ice*", and word combinations that do not have a literal meaning "such as *move heaven and earth*", are oftentimes not included in the category of idiom.

1.3. Understanding idioms

The different models that have developed to explain the process of understanding an idiomatic expression differ in the degree of "literal word activation" (Van Ginkel and Dijkstra 2020:131). They can be divided into three categories: models where "(1) figurative meaning takes precedence", approaches where "(2) literal word and/or sentence meaning takes precedence", and "(3) hybrid models where figurative and literal meaning are allowed to interact" (Van Ginkel and Dijkstra 2020:132). These categories are based on the "speed of retrieval", where "precedency for figurative meaning means that the figurative reading of an idiom becomes available at an earlier point in time than its literal sentence reading (or literal word meaning is suppressed by the idiom's figurative meaning)" (Van Ginkel and Dijkstra 2020:132).

1.3.1. Category 1: Traditional view

In the first category, we find Chomsky (1980) who understands idiomatic expressions to be non-compositional combinations of words whose figurative meanings are not associated with the literal meanings of the separate words (Cieślicka 2006:117, Gibbs and Nayak 1989:100). Alternatively, in Hamblin and Gibb's (1999:26) words, idioms are "frozen, semantic units that are essentially noncompositional". This means that speakers will fail to understand the meaning of an idiom, for instance *wear the trousers*, through analysis of the separate parts *wear, the,* and *trousers* (Hamblin and Gibbs 1999:26). The idiom's figurative meanings are listed in the mental lexicon similarly to every other individual word in the dictionary (Gibbs, Nayak and Cutting 1989:567). However, unlike understanding literal language, idiom comprehension functions either through direct retrieval from the lexicon, once the literal meanings have been judged to be unsuitable, or at the same time as the processing of the literal meanings (Gibbs, Nayak and Cutting 1989:567).

Chomsky's approach is consistent with the traditional perspective on the understanding of figurative language, also known as the "standard pragmatic model", which perceives figurative language as atypical compared to "'normal' literal speech" (Cieślicka 2006:117; Polcar 2003:15). In this model, the language user can only comprehend the meaning of a figurative utterance after having failed to analyse the direct meaning first (Polcar 2003:10). Only when the direct meaning is discovered to not cover the full context, the listener continues to access the meaning of "an indirect speech act" or of a "different (non-literal) meaning" (Polcar 2003:10, 15). It follows that this "literally-first processing model, or standard pragmatic model," claims that the literal and the figurative system of language comprehension are two different processes and that the computing of a literal utterance precedes that of a figurative one

(Cieślicka 2006:117, Polcar 2003:15). In other words, failing of the literal analysis triggers that of the figurative processing (Cieślicka 2006:117).

The Lexical Representation Hypothesis by Swinney and Cutter (1979) is the most influential model from this first category (Van Ginkel and Dijkstra 2020:132). Their model proposes that idioms undergo the same storage and retrieval process as normal words (Swinney and Cutter 1979:525). They theorize that the "computation" of both the idiomatic and the literal meaning of the idiom string is initiated simultaneously with the first word of the string being the start signal (Swinney and Cutter 1979:525). In other words, the separate words of the string are accessed and analysed at the same time as the complete idiom phrase is being scrutinised (Swinney and Cutter 1979:525). They (1979:533) claim that idiom processing does not have "some sort of "special" process", but that it would show to have "very "normal" processing solutions".

Several arguments can be found against this approach to idiom processing. Gibbs, Nayak, and Cutting (1989:576) argue that this non-compositional view is not adaptable to the real world. They claim that many idioms are "decomposable or analyzable with the meaning of their parts contributing independently to their overall figurative meaning" (1989:576). Furthermore, the analysability of the idiom functions along a scale or a "continuum" (Gibbs, Nayak and Cutting 1989:577). Moreover, Van Ginkel and Dijkstra (2020:132) add that the time it takes for the retrieval of an idiom is less than it takes for the literal meaning of the words to be calculated. This means that the interpretation of the idiomatic phrase will become available more quickly than the literal one (Van Ginkel and Dijkstra 2020:132).

1.3.2. Category 2: Compositional models

It follows that in addition to "the traditional, non-compositional view of idioms", a group of compositional models has developed that argue that the meaning of idiomatic expressions is composed of both the literal meanings of the parts of the idiom and the context-applicable figurative meaning of the idiom (Cieślicka 2006:117). These compositional models differ in the degree of importance that is designated to the literal and figurative meaning of the idiomatic expressions during the idiom processing (Cieślicka 2006:117). Hamblin and Gibbs (1999:27) indicate that various studies from the late 80's and 90's have demonstrated that many idioms are "at least partly analyzable or decomposable". Indeed, they seem to exist "on a continuum of analyzability or decomposability" with some idiom phrases being easily analysable, such as "*pop the question"*, and others being "far less decomposable or nondecomposable, such as "*kick the bucket"* (Hamblin and Gibbs 1999:27).

In their "idiom decomposition hypothesis", Gibbs and Nayak (1989:104) explore the view that "idioms are partially analyzable and speakers' assumptions about how the meaning of the parts contribute to the figurative meanings of the whole determines the syntactic behavior of idioms". The results of their research indicate that idioms containing semantic elements that provide a contribution to the general figurative meaning such as *"go out on a limb"* will be evaluated as more "syntactically flexible or productive" than non-decomposable idioms such as *"kick the bucket"* (Gibbs and Nayak 1989:100). Thus, Gibbs and Nayak (1989:100) claim that this suggests that idioms are not part of a separate linguistic class, but that they share compositional characteristics with "more 'literal'" language. Furthermore, they state that idiom processing is first done in a compositional manner, meaning that people start by analysing the separate parts of the idiom (Cieślicka 2006:117). Gibbs and Nayak (1989:131) argue that people hearing an idiom will interpret its individual elements separately and thus will easily understand decomposable idioms. They (1989:131) base these claims on their results, showing that listeners do not find it difficult to understand decomposable idioms that have undergone a syntactical or lexical transformation such as "*[t]he law was laid down by John"* instead of *"John laid down the law"* and *"hit the sack"* instead of the original *"hit the hay".*

In addition, Gibbs, Nayak and Cutting's experiments (1989:576) indicate that decomposable idioms are processed faster than non-decomposable idioms. When people encounter *lay down the law,* they know that each element is meaningful: *lay down* meaning "the act of invoking" and *the law* referring to the rules of society (Gibbs, Nayak and Cutting 1989:587). The difference in processing time for non-decomposable idioms suggests that people usually commence a compositional analysis of the phrase in order to determine its meaning (Gibbs, Nayak and Cutting 1989:587). This method is more suited to discover the nonliteral interpretation of an idiom; thus, when a nondecomposable idiom is discovered, its meaning must be retrieved from the mental lexicon (Gibbs, Nayak and Cutting 1989:588). Moreover, Gibbs, Nayak and Cutting (1989:591) also show that there is a processing advantage for idiomatic language compared to literal language. This is due to people starting a compositional analysis, whereby the familiarity of the different idiomatic word combinations causes this analysis to be quicker than that of newer, literal strings (Gibbs, Nayak and Cutting 1989:591).

Research by Tabossi, Fanari and Wolf (2009), however, does not support the idiom decomposition hypothesis of Gibbs, Nayak and Cutting (1989). In their study, Tabossi, Fanari and Wolf (2009:529) use a "semantic judgment paradigm", where the participants determine whether a phrase has meaning or not in order to test the three principal hypotheses of idiom recognition, one of which being the idiom decomposition hypothesis (Tabossi, Fanari, and Wolf 2009:531). In the task, "the nondecomposable idioms showed the same [processing] advantage as did decomposable idioms and clichés" (Tabossi, Fanari, and Wolf 2009:534). Furthermore, they (2009:534) indicate that several studies have not managed to replicate the findings that Gibbs, Nayak and Cutting (1989) presented.

1.3.3. Category 3: Hybrid approach

The final category contains hybrid models which combine elements from both the non-compositional and the compositional approaches (Cieślicka 2006:119, Libben and Titone 2008:1103). This means that idioms may be retrieved directly when they are either familiar or easy to predict, while they may also undergo a compositional analysis during processing, particularly when the idiom is unfamiliar or difficult to predict (Libben and Titone 2008:1103).

A well-known model from this approach is the configuration model by Cacciari and Tabossi (1988), which stresses the function of literal meanings in building the interpretation of figurative idiomatic expression (Cieślicka 2006:118). Cacciari and Tabossi (1988:674, 677) find that with predictable idioms, where access to the last word is mandatory to recognize the phrase as being an idiom, only activation of the idiomatic meaning occurs, while with non-predictable idioms only the literal meaning is accessible. Furthermore, when no cues are available, activating of the idiomatic expression takes place 300 ms after the idiom has ended (Cacciari and Tabossi 1988:677). Therefore, Cacciari and Tabossi (1988:668) conclude, a new hypothesis is needed to describe "the representation and the processing of idioms". Their model proposes that "the language comprehension device" processes the idiom in a literal way up until the point where the expression is recognized as an idiomatic phrase; the "idiomatic key" (Cieślicka 2006:118; Sprenger, Levelt and Kempen 2006:163). This key is defined by Tabossi and Zardon (1995:275) as "the information in the string that has to be processed literally before the figurative meaning of an idiom can be activated". This key content word dictates the activation of the figurative process; some idioms will have this at the start of the phrase, while others require processing of the entire expression before this happens (Cieślicka 2006:118). Here, the familiarity with an idiom is the key to the quick capture of the expression instead of the idiomaticity (Van Ginkel and Dijkstra 2020:132).

It should be noted that both Cacciari and Tabossi (1988) and Tabossi and Zardon (1995) remark that this is only the case as long as there is no "biasing context" that would influence the reader for either a literal or a figurative meaning (Siyanova-Chanturia, Conklin, and Schmitt 2011:253). Indeed, Siyanova-Chanturia, Conklin, and Schmitt's (2011:254) eye-tracking study which included stories with the following three possibilities: “an idiom used figuratively (at the end of the day – ‘eventually’, an idiom used literally (at the end of the day – ‘in the evening’), or a novel phrase (at the end of the war)” shows that the figurative uses of idioms are processed with an advantage compared to the literal uses, thereby suggesting that the context of these stories sufficed to resolve any uncertainty of meaning during the processing phase (2011:265). However, the existing theories do not account for the consequences of a “biasing context”; therefore, Siyanova-Chanturia, Conklin, and Schmitt’s (2011:265) argue that they cannot accurately compare their research to these studies and that further research is needed.

1.4. Non-native speakers

Idioms often give native, or L1, speakers a processing advantage over literal language; conversely, they can be quite a challenge for non-native, or L2, speakers (Van Ginkel and Dijkstra 2019:131). Even though different theories have been proposed over the years, generally it is agreed upon that native speakers process idioms more quickly than literal language (Siyanova-Chanturia, Conklin, and Schmitt 2011:253). Swinney and Cutler (1979:528) for instance show that access to the “lexicalized idiom” for native speakers is quicker than the access to the meaning of the several words in the control phrase. Furthermore, a more recent study by Tabossi, Fanari, and Wolf (2009:533) shows evidence that all the formulaic expressions in their study, idioms and clichés, were processed faster than the control phrases. This is called the “idiom superiority effect” (Tabossi, Fanari, and Wolf 2009:533; Van Ginkel and Dijkstra 2019:131). However, L2 speakers find idioms somewhat more difficult to recognize and understand (Beck and Weber 2016:1). Even though idioms are fixed expressions, which might imply easy learning and recounting, the difference between an idiom’s figurative and literal meaning remains challenging (Beck and Weber 2016:1). A speaker claiming that they are “*in hot water”* will be understood immediately by a native speaker, while possibly confusing a non-native listener about what is happening (Beck and Weber 2016:2).

Siyanova-Chanturia, Conklin, and Schmitt’s (2011:251) eye-tracking study supports these ideas. The results of their non-native, highly proficient, participants indicate that idioms and novel strings have a similar processing speed (Siyanova-Chanturia, Conklin, and Schmitt 2011:251). In comparison, their (2011:264) native participants processed the idioms quicker than the novel phrases. Furthermore, figurative phrases have a longer processing time than literal strings (Siyanova-Chanturia, Conklin, and Schmitt 2011:251). In their “full idiom analysis”, where they (2011:258) analysed both the “reading times” and the “fixations”, Siyanova-Chanturia, Conklin, and Schmitt (2011:254) show that L2 speakers “slow down when reading idioms' figurative meanings". Therefore, they found evidence that suggests that even non-native speakers with a high proficiency in their target language may find processing of figurative idioms difficult (2011:260). In other words, "the idiom's figurative meaning incurs a processing cost when compared to its literal equivalent" (Siyanova-Chanturia, Conklin, and Schmitt 2011:263).

Similarly to native speaker research, there is no consensus on the understanding of idioms for non-native speakers. Online research that exists on the comprehension of bilingual idioms has produced contradictory results (Van Ginkel and Dijkstra 2020:131). In some studies, in bilinguals literal processing is shown to take precedence over figurative processing such as Cieślicka (2006), while in other research findings are presented that both the literal and the figurative meaning are accessible online (e.g. Beck and Weber 2016).

Cieślicka (2006:120) bases their "L2 idiom comprehension model" on the assumption that L2 speakers learn the literal meanings of words before they come across fixed expressions containing their figurative meanings. Thus, Cieślicka (2006:12) reasons that these literal meanings will receive a "more salient status" than the figurative ones during the processing of idioms by L2 learners. "Salient meanings" are those meanings that enjoy the first and the strongest activation during this process due to them being "more strongly encoded" or in other words being stored the longest and being represented in the most complete form compared to other meanings (Cieślicka 2006:121). Learners of a new language usually meet new idioms when they have already encountered the literal meanings of the separate words of the idiomatic expression; therefore, the literal meaning will have been established by that time (Cieślicka 2006:121). Cieślicka (2006:140) argues that their experimental study supports the view of the priority of literal meanings during L2 idiom processing which is in accordance with the findings of Siyanova-Chanturia, Conklin, and Schmitt (2011:266) who show that "in non-native speakers, figurative meanings required more re-reading and re-analysis than literal ones".

However, Siyanova-Chanturia, Conklin, and Schmitt (2011:254) also criticise Cieślicka's (2006) study. After listening to a sentence containing an idiom, for instance "*George wanted to <u>bury the hatchet</u> soon after Susan left*", the non-native participants were asked to pick one of four options, the following are applicable to the example mentioned: "a word related to the idiom's figurative meaning (*forgive*), its control (*gesture*), a world related to its literal meaning (*axe*), or its control (*ace)*" (Siyanova-Chanturia, Conklin, and Schmitt 2011:254). Cieślicka (2006) argues that because the targets that are connected to the literal meaning received quicker responses compared to those with the figurative meanings, the literal meanings are activated before the figurative ones (Siyanova-Chanturia, Conklin, and Schmitt 2011:254). Siyanova-Chanturia, Conklin, and Schmitt (2011:254) counter that there is a strong semantic relation between the word *hatchet* and the word *axe*; therefore, it does not come as a surprise that the literal option was preferred.

Unlike Cieślicka (2006), Beck and Weber's (2016:12) study finds that both the group of native speakers and the group of language learners display access to the figurative as well as to the literal meaning of the idiom in a non-biasing context. Their (2016:5) participants, self-assessed "skilled speakers of English" and native speakers, took part in a "lexical decision task" in which they "responded to English target words preceded by English idioms embedded in non-biasing prime sentences" (2016:1). Beck and Weber (2016:5), for instance, used the idiom "*to pull my leg*, the prime sentence "*John likes to pull my leg*", "target word WALK", and "unrelated control word" "MILK". The participants were asked to listen to sentences, followed by an English word or "non-word" on a computer screen (Beck and Weber 2016:6). They were instructed to decide whether this was English or not by selecting the "YES" or "NO" button (Beck and Weber 2016:6). The results show no significant difference between the manner in which the idioms are processed (Beck and Weber 2016:11). Furthermore, the results do not show that the translatability of the idioms from the native language to the target language has a significant influence (Beck and Weber 2016:12). They (2016:12) thus interpret these findings as evidence that language learners with a high proficiency in their L2 will process figurative meanings in a manner that is similar to native speakers. At the start of their language learning journey, L2 speakers may be slower in processing figurative language; however, as their proficiency grows, the differences between them and native speakers can be explained by general L1 and L2 language characteristics rather than differences in processing (Beck and Weber 2016:12).

These results are challenged by Siyanova-Chanturia, Conklin, and Schmitt (2011:265) who argue that their findings indicate that "idioms are not represented in the mental lexicon of a non-native speaker in the same way they are represented in the lexicon of a native speaker". Their (2011:265) research shows that the non-native participants processed the idioms and the novel strings in a comparable manner, contrary to their native participants. The processing speed of both idioms and novel strings was similar, as well the "figurative vs. novel or literal vs. novel comparisons" (Siyanova-Chanturia, Conklin, and Schmitt 2011:265).

Despite the disagreements in theories, it is evident that mastering idioms in the target language can only benefit non-native speakers (Beck and Weber 2016:2). The proficiency of the L2 would increase significantly by making the speakers sound more similar to natives (Boers et al. 2006:245). Indeed, the Common European framework of reference uses the language learners' skills of producing and understanding idioms as one of their assessment points to determine the language level for non-natives (Beck and Weber 2016:2).

1.5. Linguistic relativity

The Sapir-Whorf hypothesis, "also known as the linguistic relativity hypothesis", proposes that the language that is spoken by an individual will influence the way that they think about their reality (Lucy 2001:903). Edward Sapir and Benjamin Lee Whorf are associated with this theory having formulated the most influential thoughts on the topic of linguistic relativity in the second quarter of the last century (Lucy 2001:903). According to Lucy (2001:903), two claims must be elaborated on to argue for this hypothesis. Firstly, "languages differ significantly in their *interpretations* of experienced reality – both what they select for representation and how they arrange it" (Lucy 2001:903). Secondly, "language interpretations have *influences* on thought about reality more generally – whether at the individual or cultural level" (Lucy 2001:903).

Lucy (2001:904) interprets this hypothesis as "linguistic structures [that] influence habitual thought by serving as a guide to the interpretation of experience". Language users generally see the categorization methods and features of their language as something natural and influenced by their external reality; hence, they employ it as a "guide" (Lucy 2001:904). When speakers try to find an interpretation for an experience from one of the categories that their particular language offers, they unknowingly activate other meanings belonging to this same category that are specific to their language (Lucy 2001:904). Because language is such an omnipresent and covert element of behaviour, language users do not notice that the various associations they "'see'" actually come from their language rather than from external reality (Lucy 2001:904). Furthermore, due to lack of another language, which could either be natural or artificial, speakers do not have the ability to talk about experience in another way; therefore, the possibility of recognizing the "conventional nature of their linguistically based understandings" is lost (Lucy 2001:904). In other words, "our linguistically determined thought world not only collaborates with our cultural idols and ideals, but engages even out unconscious personal reactions in its patterns and gives them certain typical characters" (Whorf 1939:198). Thus, this system of different categories that every single language offers to its users is not a universal one, but a specific "'fashion of speaking'" (Lucy 2001:904).

Boroditsky (2000; 2001; 2018) subscribes to the theory of linguistic relativity, albeit not to the strong Whorfian sense of linguistic determinism. The view that claims that "thought and action are entirely determined by language" has lost its popularity in the field, Boroditsky (2001:2) explains. Despite the overwhelming proof that this strong view of linguistic determinism is not valid, less radical theories based on these ideas are interesting to entertain (Boroditsky 2001:2). Boroditsky's (2000; 2001; 2018) own research for instance focusses on abstract language such as temporal and spatial metaphors. Through experiments, Boroditsky (2000; 2001) argues that native language has a strong influence on the way people think about abstract concepts such as time. Evidence from cross-linguistic research shows that if the "spatiotemporal metaphors" in which people talk about time differ, their conceptions of time will also vary (Boroditsky 2000:26; 2018:651). It would seem that "metaphorical mappings" from domains that are grounded in experience, such as space, are indeed shaping the more abstract domains, such as time (Boroditsky 2000:26). Thus, when these metaphors are not the same, people's idea of time will be different (Boroditsky 2001:20). For instance, English speakers describe time as being a horizontal concept, while Mandarin speakers describe it as vertical (Boroditsky 2001:1). When the information people receive from their senses is insufficient or incomplete "as with the direction of motion of time", languages possibly perform an important part in moulding the manner in which their users think (Boroditsky 2001:20). Additionally, learning to talk about a domain in a new way, as is the case when learning a new language, can alter the way people think about that particular domain (Boroditsky 2001:19). Hence, "[l]anguage can be a powerful tool for shaping abstract thought" (Boroditsky 2001:20).

1.6. Language and the brain

Rippon (2019:109) calls the understanding of the plasticity and the mouldability of our brains one of the most important findings of the past thirty years: our brains do not just change during early development, but they continue to do so during our lives, reflecting both our experiences and, seemingly paradoxical, those we do not have. This is a departure from the 1990's where media coverage focussed on "critical periods of brain development" (Thompson and Nelson 2001:6), where the focus was put on the failure to create "core competencies" if the specific time window for the input had passed (Rippon 2019:110). Even though it was understood that children's brains had a certain level of redundancy, the assumption was made that the "structures and connections in the brain were hard-wired, fixed and unchangeable" (Rippon 2019:110). During this period, "[b]iology was most definitely destiny" (Rippon 2019:11). Thompson and Nelson (2001:7) call these contributions "valuable"; however, they caution that the media back then also had the tendency to exaggerate the then status of neuroscience by overinterpreting results and overpromising its uses. Furthermore, emphasising "critical brain development […] also risks ignoring important achievements of later years, as well as the enduring plasticity of the mature brain" (Thompson and Nelson 2001:10). The discovery that "[t]he brain preserves the ability of experience-dependent plasticity throughout life" (Schmidt et al. 2020:1) shows that it is important to take the impact that the outside world has on our brains into account (Rippon 2019:110). Therefore, the discussion whether our brains are influenced by either nature or nurture must change to research the degree of entanglement "of the 'nature' of our brains" with the "brain-changing 'nurture' provided by our life experiences" (Rippon 2019:110).

The findings that our brains are capable of lifelong plasticity can offer insights into the changes that are initiated through experiences in real life; furthermore, it follows that we need to be more aware of what exactly is present in the world as our brains will be altered through engagement with it (Rippon 2019:113). Every day, our brains are constantly monitoring the predictions they make based on large amounts of input and the real outcomes; this data is analysed and updated (Rippon 2019:113). The main goal of this system is to "minimise 'prediction error'" in order to be as efficient as possible and to reduce the need to double check or overthink (Rippon 2019:113). Tamir and Thornton (2018:206) confirm that the "brain is well suited to make such predictions"; indeed, the "theory of predictive coding" proposes that "the brain represents the world not as percepts, but as predictions". People make predictions reflexively in the domain of perception for instance by catching a ball in a game or using contextual signals to predict that there will be a ball in a football stadium (Tamir and Thornton 2018:206). Furthermore, they will make predictions in the domain of cognition for example by finishing someone's sentence (Tamir and Thornton 2018:206). Finally, this same process may also happen in "the domain of social cognition": people can predict the actions of the other people around them based on their identities and their previous experiences with them (Tamir and Thornton 2018:206).

Accordingly, our brains are efficient; however, their predictions are not faultless (Rippon 2019:115). The system's templates like to generalize information in order to be come to conclusions quickly; they are "the ultimate stereotypers" using personal experiences, cultural and social rules, and predictions based on the surroundings (Rippon 2019:116). Rippon (2019:117) offers the example of "deep learning" to clarify our behaviour: "computers are programmed to extract patters from information and to 'self-train' […] rather than be programmed to carry out specific tasks". Neuroscientists have found that if the data that is being put into the system is "intrinsically biased", the system will acquire a biased rule (Rippon 2013:117). For instance, Bolukbasi et al. (2016:1) discovered that word embeddings, "trained only on word co-occurrence in text corpora", would finish the statement "man is to computer programmer as woman is X" with the word "homemaker". Consequently, our brains can be seen as such a system; extracting information from the world and establishing rules based on this input (Rippon 2019:118).

In addition to handling everything that we want, need, believe, or desire, we have to manage anticipating those of the people around us oftentimes based on unspoken cultural and social rules (Rippon 2019:120). Tamir and Thornton (2018:201) focus on the two main challenges of social cognition: "organizing social knowledge efficiently and using it for social prediction". They argue that "the organizational dimensions of social content may scaffold social prediction" (Tamir and Thornton 2018:201). In other words, we will learn scripts that encompass the rules of our respective societies and we will use those to make predictions, allowing us to display the correct social behaviour (Rippon 2019:121). Stereotypes will be found in these scripts as they are "social shortcuts" that help us react quickly; however, this does not necessarily mean that they are accurate as was shown in the previous paragraph (Rippon 2019:121).

Allport (1954:191) defines stereotypes as "exaggerated belief[s] associated with a category; Quadflieg and Macrae (2011:2016-2017) largely agree, but contemporize the definition as "cognitive structures that provide knowledge, beliefs, and expectations about individuals based on their social group membership". It can be understood that stereotypes have a purpose – creating shortcuts in our brains that help us navigate the world around us; however, these stereotypes also incorporate the expectations of your own self such as how to behave as a male or female, who to socialize with, and what (career) path to follow while growing up (Rippon 2019:121, 356). Stereotypes can reinforce themselves, either because they happen to be true and the girls are sitting quietly in a corner reading and the boys are loudly playing a ball sport outside, or because they include "an element of self-fulfilment" and the girls fail a science test after repeatedly hearing that women are bad at science (Rippon 2019:356).

Quadflieg and Macrae (2011:229-230) go even further and argue that as stereotypes are multifaceted "(i.e., constituted by several different facets which also differ according to their salient properties)", different neural networks may add their contributions to those various elements of the stereotype. For instance, gender stereotyping can take several forms; it focusses on gendered clothing, activities, and personalities (Quadflieg and Macrae 2011:230). Based on findings that conceptual knowledge is distributed across "networks of modality-specific cortical areas", Quadflieg and Macrae (2011:23) assert that this leads to activity in the relevant neural network. This leads Rippon to demand from us that we continue challenging gender stereotypes: "[a]nd as we know, our brains will reflect this input. Stereotypes could be straightjacketing our flexible, plastic brains. So, yes, challenging them does matter" (2019:356-357).

1.7. Linguistics and feminism

One way of challenging these stereotypes is by researching sexist use of language. This has been for some time, and still is, a "grassroots-based activity of feminists" who are interested in language and the way the sexes are represented in it (Pauwels 2003:552). Their research reveals the similarities in which women and men are portrayed in different speech communities and languages (Pauwels 2003:552). One of the most noteworthy of these is the "*asymmetrical treatment*" where the man is seen as the "prototype for human representation", while the woman is reduced to the "invisible" or the "marked" one (Pauwels 2003:552). A process that Criado Perez (2019:3) calls "[t]he male-unless-otherwise-indicated approach". In other words, women need to be actively made visible in the language through linguistic processes such as derivations of the male construction (Pauwels 2003:552). Another typical asymmetry can be seen in the lexicon of a language; the ""male as norm" principle" shows through gaps in vocabulary where words to describe women in certain "roles, professions, and occupations" are missing (Pauwels 2003:553). Conversely, there are areas that are dominated by women where specific nouns to denote men are absent (Pauwels 2003:552). Especially this "semantic asymmetry" is investigated by feminist linguists as they are an indication of the perceived worth of the sexes in society (Pauwels 2003:552).

It would perhaps seem logical to think that inequality is part of language due to historical developments; however, even new languages show bias (Criado Perez 2019:8). A 2015 report on emoji found that 92% of people online use them; more specifically, 78% of women and 60% of men are frequent emoji users (*Shaul* 2015). However, the emoji language was predominately male until 2016 (Criado Perez 2019:8). Unicode, a "Silicon Valley-based group of organizations that work together to ensure universal, international software standards", did not specify the gender of most emojis (Criado Perez 2019:8). Thus, the emoji of a running man was called runner without gender specification and the platforms on which the emoji was used then interpreted it as being male (Criado Perez 2019:8). In 2016, Unicode changed this by explicitly gendering their human emojis (Criado Perez 2019:8). Criado Perez (2019:8) calls this a "small victory, but a significant one".

Similarly to Criado Perez, most linguistic activists advocate for language change to achieve the goal of a more equal representation of the sexes in the language (Pauwels 2003:552). Many of them follow the "interactionist view of language and reality" which is based upon a less strict version of the Sapir-Whorf hypothesis: "language shapes thought and reflects social reality" (Pauwels 2003:554). There are many views on why language change should happen; the following three reasons are the main incentives: "(1) a desire to expose the sexist nature of the current language system; (2) a desire to create a language which can express reality from a women's perspective; or (3) a desire to amend the present language system to achieve a symmetrical and equitable representation of women and men" (Pauwels 2003:555).

Pauwels (2003:561) concludes that in language communities that speak English as well as some with a mainly European language, awareness of the existing gender bias has increased due to the engagement of feminist activists. Even though there are still many language users who claim that there is no bias or who continue to use sexist language, they are nonetheless aware of these problems (Pauwels 2003:561). Indeed, people start to increasingly show metalinguistic conduct which indicates a growing awareness of gender bias in language (Pauwels 2003:561). This is especially necessary, because "in the mouths of sexists, language can always be sexist" (Cameron 1985:90). However, Pauwels (2003:567) remains optimistic based on the results of linguistic activism which have contributed towards "exposing the ideologization of linguistic meanings to the speech community at large and […] challenging the hegemony of the meanings promoted and authorized by the dominant group or culture, in this case men".

1.8. Proverbs and idioms: A source of sexism

Research of Kim et al. (2019), Lomotey (2019) and Schippers (2010) are examples of investigations of sexism in language that challenge the status quo. Furthermore, their studies are specifically directed at the area of proverbs and idioms. The research of this thesis hopes to expand further on these foundations.

Kim et al. (2019:26-27) deduct from previous research that literature is seen as a "fundamental socializing agent" that reflects the norms and values in our social and cultural environment and thereby influences the way its readers think and behave. Through reading literature, children acquire the culturally appropriate assumptions, attitudes, and behaviours, and learn about the characteristics that are associated with a specific gender (Kim et al. 2019:27). Proverbs, which Kim et al. (2019:27) see as a "simplified form of folk literature passed from one generation to another", are treasured as a collection of "wisdom, knowledge, advice, and guidance, and as tools to teach cultural values and norms to members of society" in most traditional countries. Kim et al. (2019:27) introduce the example of the Malaysian "*pantun*", a type of classic Malay poetry, "*peribahasa* and *simpulan bahasa"*: these are proverbs and idioms that are repeated and memorized. Thus, Kim et al. (2019:27) argue that these portray a "microcosm" of conventional communication and an accurate depiction of the Malay state of mind".

Indeed, Kim et al.'s (2019:37) analysis of Malay and Korean proverbs and idioms shows the overwhelming influence these proverbs and idioms have had on the degree to which the patriarchy has been supported and strengthened through their use. Furthermore, they (2019:37) conclude that the 16,000 proverbs and idioms that they have used for their research display their function as "mechanisms that (i) enforce the ideal image and behavior of women, (ii) reproduce stereotypical views of the society on the nature and role of women, and (iii) deter behaviors that are deemed unacceptable by society". These include opinions on women that are established through the dominant patriarchal viewpoint; moreover, these stereotypes have been used, and are still being used, to further emphasise and maintain patriarchal patterns (Kim et al. 2019:37). Therefore, Kim et al. (2019:37) conclude that proverbs and idioms are a key to the understanding of gender equality in this region.

During their long-term research project culminating in a database of 16,000 proverbs from 240 different countries, Schipper (2010:18) discovered that the themes of proverbs and their variations that exist in one "culture, country, language or area" can often also be found in other areas. This database of proverbs reveals that "cross-culturally, different metaphors frequently convey the same message" and "even exactly the same images occur across cultures" (Schipper 2010:299). Schipper (2010:18) finds that this is due to the fixed patterns that exist regarding the distribution of status and labour in society, although things are currently changing:

> "In virtually all societies men fare better than women. Men exercise more power, have more status and enjoy more freedom. Men usually head the family, exercise considerably more force in legal, political, and religious matters, take alternative sexual partners, may often take more than one wife, have greater freedom in the choice of a spouse, usually reside near their own kin, and have easier access to alcoholic beverages and drugs. Women, on the other hand, are often segregated or avoided during menstruation, must often share their husbands with one or more co-wives, are blamed for childlessness, and are often forced to defer to men in public places. Child rearing is the only domain where women regularly exert more influence than men" (Levinson and Malone 1980:267 cited in Schippers 2010:18).

These proverbs, therefore, are indispensable in gender studies. Schipper (2010:20), in agreement with Kim et al. (2019), argues that proverbs regarding women can be used to explain the "growing gap" that has caused men and women to refrain from "sharing both public roles in life and responsibilities at home". Through the repetition of proverbs advocating the maintenance of this gendered gap based on "relatively insignificant body differences", proverbs have strengthened the existing hierarchies and built strict models of what it means to be a woman, thereby authenticating the roles for both sexes (Schipper 2010:20). This prescriptive way of looking at gender roles also causes people who do not fit the mould to be stigmatized, by both other women and men (Schipper 2010:20). "Privileges are never given up easily" and sharing the least favourite tasks is an offer that is rarely made from one who is in the dominant, and comfortable, position (Schipper 2010:20). "A bad home sends you for water and firewood" says the Rwandan proverb, thereby warning the man that he would turn into his wife's slave if he would stoop to do "women's work" (Schipper 2010:20). The European equivalent is the well-known proverb, or idiom as it is classified in this study, complaining of women "wearing the trousers" (Schippers 2010:20). These warnings for men, combined with rules and regulations for women, show "uncertainty as well as fear for loss of the status quo" according to Schippers (2010:20). Finally, Schippers (2010:20) makes the astute observation: "[i]f women had been as subservient as they ought to be according to the endless series of prescriptions and proscriptions, the proverbs would have been completely superfluous, would they not?"

In the Western world, proverbs are less current than advertisement slogans (Schipper 2010:21). However, they are still used during everyday communication, in the media, and even in political speeches (Schipper 2010:21). It is evident that a substantial number of ideas that are brought up in proverbs are no longer as current as they must have been in the past; this leads to assume that traditions are evolving, particularly in industrialized countries (Schipper 2010:19). These developments are mainly advantageous to "privileged groups" as opposed to the majority that do not have the same educational possibilities and thus persist in their proverbial mindset about what male and female is (Schipper 2010:19). However, it is certain that these proverbial ideas have also not vanished entirely amongst the "socially privileged groups"; they are oftentimes present as an "internalized subconscious legacy" (Schipper 2010:20). This is supported by Lomotey's (2019:161) research who finds that proverbs that are overly sexual are outdated in peninsular Spanish; however, the "androcentrism and sexism" are still present in contemporary proverbs due to the refashioning of the old proverbs into new forms that are now outwardly benevolent to women, but remain covertly sexist.

2. Research methodology

This thesis will use two methods of quantitative research. The first will use a questionnaire as the means of data collection which will be analysed with statistics, whereas the second method is comprised of data collection from the Corpus of Historical American English (COHA) which will be examined with non-statistical procedures. The idioms that feature in both the questionnaire and the COHA search are classified based on the Ambivalent Sexism Theory by Glick and Fiske (1996, 2010). The results will allow a comparison between the predicted bias in the answers from the questionnaire and the results from the COHA. Furthermore, it will also provide an opportunity to compare the state of benevolent and hostile sexism in the language. Finally, the difference in results will also give information about the status of native and non-native speakers concerning idioms and their internal sexism.

2.1. Ambivalent Sexism Theory

Sexism is a prejudice; however, Glick and Fiske (1996:491) argue that it is a special instance of prejudice that is characterized by a "deep ambivalence, rather than a uniform antipathy toward women". Typically, sexism is defined as "a reflection of hostility toward women"; nevertheless, this point of view ignores a certain part of sexism: the positive sentiments that are subjectively felt towards women that are frequently combined with sexist hostility (Glick and Fiske 1996:491). Allowing for this observation, Glick and Fiske (1996:491) see sexism as a multidimensional concept containing two categories: "hostile and benevolent sexism"; melding two previously separate research areas together (Lee, Fiske and Glick 2010:395). The meaning of the first area is in the name and follows Allport's (1954:9) definition of prejudice: "an antipathy based upon a faulty and inflexible generalization". The second is defined as "a set of interrelated attitudes toward women that are sexist in terms of viewing women stereotypically and in restricted roles but that are subjectively positive in feeling tone (for the perceiver) and also tend to elicit behaviors typically categorized as prosocial (e.g. helping) or intimacy-seeking (e.g. self-disclosure)" (Glick and Fiske 1996:491). Glick and Fiske (1996:491-492) caution to refrain from thinking of benevolent sexism as something good; despite the positive sentiments it can express to the perceiver, these are based on "traditional stereotyping and masculine dominance" and their repercussions are often harmful. They (1996:492) provide the example of a workplace where a male worker calls a female colleague "cute". The man might have good intentions; however, it may also give the female co-worker the impression that she is not taken seriously in her field of work (Glick and Fiske 1996:492).

Through the subjective positive feelings of the perceiver, the behaviours that are prosocial, and the efforts that are made to obtain intimacy, it shows that benevolent sexism does not fit into the standard mould of prejudice (Glick and Fiske 1996:492).

As humans, we naturally feel hostile towards groups that are different in their physical appearance; however, the fact that the difference is based on the biology of sex leads to a situation that is unlike normal group hostilities (Glick and Fiske 1996:492). In order to receive heterosexual intimacy and succeed in reproduction, humans need to depend on each other (Lee, Fiske and Glick 2010:396). Furthermore, men will also look at women for satisfaction of their requirement for psychological intimacy, because they cannot find this with other men who are essentially their competition in the race for status and resources (Glick and Fiske 1996:492). This leads to a society where the group of dominant beings, the men, must depend on the subordinate ones, the women (Lee, Fiske and Glick 2010:396). Women have "dyadic power" in this situation (Guttentag and Secord 1983 cited in Lee, Fiske and Glick 2010:396). This "dyadic power" is visible in the social ideology that advocates an attitude of protectiveness toward women, of reference toward wives and mothers, and of idealizing potential objects of romantic love (Glick and Fiske 1996:492). These attitudes are exactly those that define benevolent sexism (Glick and Fiske 1996:492).

The degree of hostility that is displayed in benevolent sexism varies across cultures (Glick and Fiske 1996:492). Although the term benevolent sexism hints at a positive opinion of women, albeit subjectively, it entails the same assumptions as hostile sexism: women belong to the domestic domain and are "the "weaker" sex" (Glick and Fiske 1996:492). Both are used as a justification of the structural power that men have by characterizing women as incompetent to have any say in economic, legal, or political areas, and by rationalizing their role in the domestic sphere (Glick and Fiske 1996:492). This way, benevolent sexism can cover up hostile sexism by claiming that "I am not exploiting women; I love, protect, and provide for them" (Glick and Fiske 1996:492).

This analysis leads Glick and Fiske (1996:493) to propose their theory that hostile as well as benevolent sexism circles around matters of "social power, gender identity, and sexuality". This is later updated to "gender hierarchy and power, gender differentiation, and heterosexuality" by Lee, Fiske and Glick (2010:396). Glick and Fiske name these three areas: "Paternalism, Gender Differentiation, and Heterosexuality", respectively, each area containing both a hostile and a benevolent element (Glick and Fiske 1996:493). First of all, paternalism clearly shows why sexism is a structure of ambivalence, because it contains a dominative side and a protective side

(Glick and Fiske 1996:493). From this point of view, women are not seen as adults and are thereby thought to be in need of a father figure; however, because men are also dependent on women as their wives, mothers, and love interests, they also see the need to love and protect them (Glick and Fiske 1996:493). Dominative paternalism thus seeks a justification of male dominance through stereotypes that depict men as strong and competent and women as the opposite of that (Lee, Fiske and Glick 2010:396). Protective paternalism only sounds positive because the women are pictured as being deserving of the protection that the men offer and they are promised rewards for their idealizing behaviour of the dominant group (Lee, Fiske and Glick 2010:396). Second, gender differentiation entertains the same tendencies as paternalism: competitive gender differentiation seeks to socially justify the structural power that men have, while complementary gender differentiation views the positive traits of said wives, mothers, and love interests as complementary to those that the man stereotypically lacks (Glick and Fiske 1996:493). The first portrays men as having the characteristics to function well in high power positions, while women are simply seen as the weaker sex (Lee, Fiske and Glick 2010:396). The latter manages to find favourable characteristics in women, though they only exist to complement those of men (Lee, Fiske and Glick 2010:396). Hence, the observation that "these depictions of women not only *complement*, but also *compliment*!" (Lee, Fiske and Glick 2010:396). This strategy further justifies the status quo of society by emphasizing that women's positive attributes such as being caring coincidentally fit well into subordinate functions such as homemaker (Lee, Fiske and Glick 2010:396). Finally, heterosexual intimacy comes from the desire for a relationship with psychological closeness; however, men's dependency on women for sex creates a resentment due to the belief that women use this alleged power to dominate men (Glick and Fiske 1996:493-494). It is clear that "heterosexual romantic relationships represent a […] double-edged sword" (Lee, Fiske and Glick 2010:396). On the one hand, men need women to form romantic relationships and to be "complete"; while on the other hand, they see women as "sexual objects" who will use this power to their benefit (Lomotey and Chachu 2020:72).

2.2. The idioms of this study

The idioms for this thesis were chosen based on the criteria from Glick and Fiske's (1996, 2010) framework. The expected bias, in the otherwise neutral-looking idioms, will be tested through a questionnaire and corpus research. Detailed information about these can be found in chapter 3 and chapter 4 respectively. The following paragraphs will elaborate on the six categories and their assorted idioms.

2.2.1. Dominative paternalism

The idioms from the category of dominative paternalism display the idea that women are incompetent and therefore need men to control them (Lomotey and Chachu 2020:72). The idioms from the following Table 1 show their subjects to be the person with the power to control other people. The subject is the figure of authority and is thus expected to be male, while the objects, or the people they are controlling, are expected to be female. In other words, the person *who wears the pants / trousers* for example, or who is in the position of power, is expected to be male as are the subjects of the other idiom searches. Even though this particular idiom is often used to refer to women who challenge male dominance, this is mostly done in a denigrating manner. The authority still lays with the male population and thus the subject must be male.

Idiom	**COHA search**
Who wears the trousers/pants	his / her trousers/pants
To sit in the driver's seat	he / she drives
To call the shots	his / her shots
To have the final say	his / her (final) say
To run the show	his / her show

Table 1: Dominative paternalism

2.2.2. Protective paternalism

The idioms representing protective paternalism indicate that men are there to protect women, because men have more strength and power in society and women depend on them for their well-being (Lomotey and Chachu 2020:72). The following idioms indicate a person who takes the responsibility to both protect and provide for another person. *To pick up the tab* fits into the last category, the other four fall into the first. It follows that all the subjects of these chosen idioms are expected to be male, while the objects are female. In other words, a male bias is predicted for these idioms.

Idiom	**COHA search**
To take someone under your wing	his / her wing
To pick up the tab	his / her tab
To stick your neck out	his / her neck
To put your head on the block	his / her head
A pillar of strength	his / her strength

Table 2: Protective paternalism

2.2.3. Competitive gender differentiation

The idioms in the category of competitive gender differentiation will show that their subjects have the necessary attributes to function in high power positions, which in turn can be used as a justification for their power (Lomotey and Chachu 2020:72). As with the previous idioms, the subjects are predicted to be male and the objects female. The object is seen as inferior and therefore not capable of having structural social power. Accordingly, the subject in these idioms is the person that displays the essential character traits and is therefore justified in its social dominance. The exception to this is obviously *to be out of someone's league* as according to this reasoning *she will be out of his league*, instead of *him being out of her league*. It follows that the corpus research is expected to deliver male bias for each of the idioms.

Idiom	**COHA search**
To be out of someone's league	his / her league
To be on the ball	his / her ball
Smart as a whip	his / her whip
Quick on the trigger	his / her trigger
To be up to snuff	his / her snuff

Table 3: Competitive gender differentiation

2.2.4. Complementary gender differentiation

The idioms of complementary gender differentiation will see the women displaying positive characteristics that fit into the traditional division of gender roles, thereby complementing the men and making up for the traits that they lack (Lomotey and Chachu 2020:72). Contrary to the previous idioms, these will therefore be expected to show female subjects. All idioms show the characteristics of a woman that belongs to the stereotype of the hard-working and obedient housewife. Therefore, the results are predicted to display a female bias. The exception is *to be at one's beck and call*; the woman in the role of the traditional wife will be at her husband's beck and call and will thus have an expected male bias with *his call* being the most frequent instead of *her call*.

Idiom	**COHA search**
To break your back	his / her back
To cook up a storm	he / she cooks
Pull your weight	his / her weight
To be at one's beck and call	his / her call
To hold one's tongue	his / her tongue

Table 4: Complementary gender differentiation

2.2.5. Heterosexual hostility

The category of heterosexual hostility contains idioms that portray a combination of sex and power (Lomotey and Chachu 2020:72). On the one hand, men tend to view women as mere "sexual objects" and though they see themselves as powerful; on the other hand, they also believe that female sexuality can be employed to take this power away from them and is thus something to be afraid of (Lomotey and Chachu 2020:72). The idioms *love 'em and leave 'em* and *wandering eye* clearly show the aspect of women being seen as sexual objects that can be used and discarded by the subject, the man, holding the power. *Of easy virtue* also belongs to this part of the category; however, this idiom is meant as a denigrating comment towards the female subject. Furthermore, *to be tied to the apron strings* refers to the female of the relationship, after having taken away the power of the man. *The old ball and chain* also refers to this part of heterosexual hostility. Therefore, *wandering eye*, *love 'em and leave 'em,* and *the old ball and chain* are expected to show a male bias, while *of easy virtue* and *to be tied to the apron strings* are expected to indicate a female bias.

Idiom	**COHA search**
Wandering eye	he / she wanders
Love 'em and leave 'em	he / she leaves
Of easy virtue	his / her virtue
The old ball and chain	his / her chain
To be tied to the apron strings	his / her apron

Table 5: Heterosexual hostility

2.2.6. Heterosexual intimacy

The idioms belonging to the last category of heterosexual intimacy refer to the idea that the woman in the romantic relationship will make the man "complete" (Lomotey and Chachu 2020:72). Thus, the focus of these idioms will be on the man getting the best deal out of a relationship, and improving their life by completing it with a woman. Therefore, the woman will *be the apple of his eye*. She will also be *his trophy*. The relationship will be *his match made in heaven*, because it will complete him. Due to the focus being on the man, he will be the one that actively *hitches,* while the woman will be the one *getting hitched.* Furthermore, the woman would need to be *the better half* to be able to complete another human being; thus, she would have to be *his better half.* Consequently, these queries will most likely result in male bias.

Idiom	**COHA search**
To be the apple of my eye	his / her eye
The better / other half	his / her half
Trophy wife / husband	his / her trophy
To get hitched	he / she hitches
Match made in heaven	his / her match

Table 6: Heterosexual intimicy

2.3. Corpus linguistics

Corpus linguistics is an "empirical discipline" that analyses data that is collected from the outside world instead of being based on the knowledge of the researchers themselves (Zufferey 2020:1). The word *corpus* means body in Latin; thus, a "*text corpus* literally *embodies* a *set of texts"* for study purposes (Zufferey 2020:1). In other words, a corpus is "a collection of spoken or written texts to be used for linguistic analysis and based on a specific set of design criteria influenced by its purpose and scope" (Weisser 2016:13). Corpora can either be synchronic "(i.e. 'contemporary')" or diachronic "(i.e. 'historical'/comparative)" and they can either be written or spoken (Weisser 2016:15).

Modern corpus linguistics does not function well without computers, because it relies on computer science to execute quantitative research based on computerized texts (Zufferey 2020:9). Specific corpus linguistic tools have been developed to support this research (Zufferey 2020:9). Especially concordances, which will primarily be used in this research, are a convenient tool to find the occurrences of a particular word together with its context (Zufferey 2020:9). It is an "analysis technique that allows linguists to investigate the occurrences and behaviour of different word forms in real-life contexts, that is, in situations where they have actually been used by native or non-native speakers" (Weisser 2016:67).

Both researchers and corpus creators have focussed on corpus size since the start of corpus linguistics (Egbert, Larson and Biber 2020:4). In the 1960's, the first electronic corpus that was created, the Brown corpus, consisted of 500 texts and "a million words of written American English" (Egbert, Larson and Biber 2020:42). Since then, corpuses have advanced tremendously; an example is the ENCOW corpus with 16 billion words (Egbert, Larson and Biber 2020:4). Throughout the years, corpus size has been an important aim in corpus linguistics (Egbert, Larson and Biber 2020:4). Indeed, many scholars have argued for a focus on size; however, there have also been voices who cautioned only looking at the size and recommended considering other aspects such as representativeness as well (Egbert, Larson and Biber 2020:4). It is evident that a bigger sized corpus is better when all the other conditions are equal; it will allow to reach higher, and more accurate, frequencies of language features (Egbert, Larson and Biber 2020:4). Furthermore, the corpus that contains more words will probably show certain *types* that a smaller corpus would not have presented (Egbert, Larson and Biber 2020:4). However, "in reality, "all things" are almost never equal", and decisions and compromises need to be made based on the specific corpus (Egbert, Larson and Biber 2020:4-5).

As corpus linguists, we aim to use a corpus that "is as representative as possible of a target population of interest" (Egbert, Larson and Biber 2020:5). To achieve this, we can either create our own corpus, or choose an existing one that fits our needs (Egbert, Larson and Biber 2020:5). The first option would only work in an ideal world without budget limitations; therefore, it is more practical to choose to use a corpus that is publicly available (Egbert, Larson and Biber 2020:5). This means that researchers try to select the corpus that is the most appropriate for their research while acknowledging the limitations it has regarding the research goals (Egbert, Larson and Biber 2020:5). Therefore, Egbert, Larson and Biber (2020:6) advice researchers to familiarize themselves with the corpus before using it in their studies. They recommend two steps that every user should follow: first, to read and examine the metadata and information that is provided about the corpus by the creators, and second, to analyse the texts that are actually contained in the corpus (Egbert, Larson and Biber 2020:6).

2.3.1. Corpus of Historical American English

For this thesis, I will be using the Corpus of historical American English (COHA), because the templates that were provided by Univ.-Prof. Mag. Dr. Ritt are specifically designed for this corpus. It would of course be possible to redesign these to fit another corpus; however, that would be a time-consuming task with the level of my current expertise that would not be feasible within the limits of an MA thesis.

The COHA is a diachronic, written corpus. It includes more than 475 million words that were taken from more than 100,000 texts from the 1820s-2010s (*COHA* 2022). The main genres are: tv/movies, fiction, magazine, newspaper, and non-fiction (*COHA* 2022). The COHA is "balanced by genre, sub-genre and domain across decades" for example the total of each decade is composed of 48-55% of the genre 'fiction' starting in the 1810's up to the 2000's (Alatrash, Schlechtweg and Schulte im Walde 2021:1-2). This gives researchers the reasonable certainty that the data shows changes that have actually occurred instead of "just being artifacts of a changing genre balance" (*COHA* 2022). Alatrash, Schlechtweg and Schulte im Walde (2021:1) even cleaned up the corpus last year to "overcome its main limitations, such as inconsistent lemmas and malformed tokens, without compromising its qualitative and distributional properties".

The main function of the COHA, and the one that will be used for this study, is the ability to scan for words in different ways (Alatrash, Schlechtweg and Schulte im Walde 2021:2). Online COHA searches are "case insensitive"; it is thus unnecessary to activate a separate query for capitalized words (Alatrash, Schlechtweg and Schulte im Walde 2021:3). Another practical option is the possibility to search "on the syntactic level as well" due to "its lemmatized and part-of-speech tagged data" (Alatrash, Schlechtweg and Schulte im Walde 2021:3). It is possible to search for certain constructions by using the part-of-speech "tag" *_n* or *_v* to search for nouns or verbs respectively. Moreover, it is possible to use the lemma of a word by writing it in uppercase in the search such as "*LADY_n* to retrieve words such as lady, ladies, Lady, Ladies" (Alatrash, Schlechtweg and Schulte im Walde 2021:3).

2.3.2. Gender bias in the COHA

The data that makes up the COHA is not free of bias; Jennejohn, Nelson and Nuñez' analysis (2021:767) finds evidence of this. First, gender bias can be observed in the use of "occupational words and adjectives" (Jennejohn, Nelson and Nuñez 2021:767). Jennejohn, Nelson and Nuñez (2021:795) are not surprised to see that the most female-biased occupations in their data set are "*nurse, housekeeper,* and midwife", while the most male-biased ones are "*postmaster, sheriff,* and *surveyor*". Expected are also the adjectives which are "*feminine, charming,* and *gentle"* for the female-biased, and "*obnoxious, unscrupulous,* and *autocratic*" for the male-biased ones (Jennejohn, Nelson and Nuñez 2021:795). This data supports the hypothesis that the authors of the texts that constitute the corpus were affected by the gender stereotypes of their time (Jennejohn, Nelson and Nuñez 2021:795). Second, the results show that gender bias generally becomes less prominent over time (Jennejohn, Nelson and Nuñez 2021:796). The earlier the decade, the more prominent the gender bias (Jennejohn, Nelson and Nuñez 2021:796). Interestingly, the bias that was measured for a specific word can also show "bias "swings"" with changes from one decade to the other (Jennejohn, Nelson and Nuñez 2021:767). While a word such as "*attractive*" has persistent collocations with female associations, and "*organized* and *tough*" with male, "*modest*" switches throughout the decades from female to neutral to male (Jennejohn, Nelson and Nuñez 2021:800). It follows that the results of such research are influenced by the decade of the COHA that is being used; therefore, any efforts to adjust for this bias will cause the results to be inaccurate (Jennejohn, Nelson and Nuñez 2021:800). Lastly, all texts show gender bias; however, those written by female authors show a less extreme form (Jennejohn, Nelson and Nuñez 2021:767). As expected, most authors of the texts in the COHA are male (Jennejohn, Nelson and Nuñez 2021:801). The female authors are outnumbered in every decade up to the 1990s (Jennejohn, Nelson and Nuñez 2021:801). However, the differences between them are not as striking as would be expected (Jennejohn, Nelson and Nuñez 2021:801). Male authors do write in a more biased way compared to female authors; nevertheless, they both produce biased texts (Jennejohn, Nelson and Nuñez 2021:801).

2.3.3. Excel

The Excel templates into which the resulting numbers from the COHA are fed contain pre-set calculations for either *he/she* + [verb] or *his/her* + [noun]. The number of occurrences of the verb or noun are filled in for every decade from 1820 to 2010 and these results are transformed into graphs that portray the gender bias of the specific queries. The following Table 7 provides an insight into the construction of the Excel template; a part of the table of *his* [noun] is given as an example here. *His* [noun], *her* [noun], *he* [verb] and *she* [verb] each have their own separate table. The occurrences that were found for the searches must be filled in manually into the respective boxes. As can be observed, each decade has its own column where the number of tokens can be added; the total words of the COHA of that particular decade are already filled in, and the tokens per million words are then calculated automatically.

his [noun]				
Decade	**1820**	**1830**	**1840**	**1850**
Number of tokens				
Total words				
Tokens / Million				

Table 7: Construction of the Excel template

Following the completion of the input table, the Excel template offers several interesting graphs displaying the results. I will be using two of them for this study: the normalized frequencies occurrences, and the difference graphs. An example of the normalized frequencies occurrences is displayed here in Figure 1. It shows the occurrence of the male tokens versus the female tokens per a million words per decade. Thus, it shows the results from the last row of the table as can be seen in the preceding Table 7. In Figure 1, it can be observed that the blue line, indicating the male tokens, never dips below that of the red line, indicating the female tokens. This is already a sign that a male bias exists; however, this must be corroborated through more data.

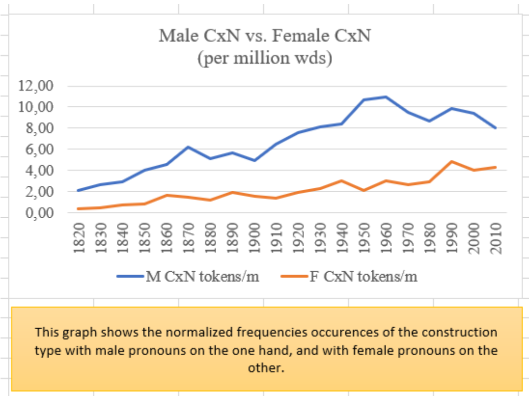

Figure 1: Example of a normalized frequencies occurrences graph

The difference graph as seen in Figure 2 displays the difference of the normalized frequencies of the male tokens minus those of the female tokens. As in the previous graph, this difference graph shows a male bias, because it does not dip below zero. A result below one would indicate a female bias, and a result of exactly one would indicate equality.

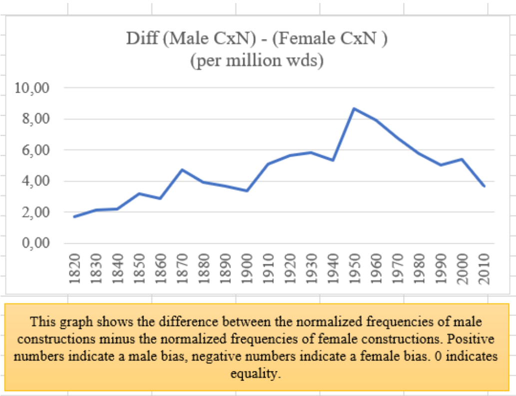

Figure 2: Example of a difference graph

In light of the findings of Jennejohn, Nelson and Nuñez (2021), it is imperative to observe the complete history of the keywords of the idioms from this study. Therefore, these graphs will be useful in nuancing the results from mere numbers to a more tangible view of the historical development of these words.

2.4. Questionnaire

Even though the term questionnaire is probably familiar, Dörnyei (2007:102) cautions that it is not that straightforward to provide a definition. First, most questionnaires are misleading in their names and do not include actual questions ending with a question mark (Dörnyei 2007:102). Second, even amongst researchers the term has been used in the more general sense of "interview schedules/guides" or "self-administered pencil-and-paper questionnaires". Thus, Dörnyei (2007) decides on the term questionnaire according to Brown's (2001:6) definition: "any written instruments that present respondents with a series of questions or statements to which they are to react either by writing out their answers or selecting among existing answers".

2.4.1. Components

The main components of a survey are: title, instructions, questionnaire items, additional information, and a final thank you (Dörnyei and Taguchi 2010:18-22). The title of the questionnaire of this study is *Master Thesis Research on Idioms*. Dörnyei and Taguchi (2010:18) indicate that a title should "identify the domain of investigation". This particular title was chosen in order to refrain from biasing the respondents by referring to anything indicating gender in the title; hence, this title announces the general domain of research, but not the specific area. The instructions to the survey follow the title (Dörnyei and Taguchi 2010:18). There are two types: "[g]eneral instruction (or "opening greeting")" at the start, and "specific instructions" that introduce every task (Dörnyei and Taguchi 2010:18). The general instruction should discuss at least the following points: the purpose of the study, the organization that conducts the study, the importance of there being "no right or wrong answers", a promise of the answers remaining confidential, and "saying "thank you"" (Dörnyei and Taguchi 2010:19). All elements are present, except the purpose of the study. Instead, there is an explanation about this lack of information:

> "I know that this does not reveal much about the actual research, but I hope you can understand that providing more information at this stage would bias the results. If you have any questions and/or you would like more information about my research, please send me an e-mail: fiona.tamminga@gmail.com; I will be happy to tell you more about the study."

After these general instructions, the main body of the survey starts with the required questionnaire items (Dörnyei and Taguchi 2010:20). These items consist of two stories portraying the characteristics of the two protagonists York and Rome, and the following Likert scale questions asking to score whether the proposed idioms are more in line with either York's or Rome's character. This central part is finished by several factual questions that inquire about the respondents themselves. Following the survey items, additional information can be found if this matches the circumstances (Dörnyei and Taguchi 2010:21). It was decided to combine this with the final element of a questionnaire that Dörnyei and Taguchi (2010:21) propose, the final "thank you". As a result, the last sentence is the following: "Thank you for your participation. If you have any questions and/or you would like more information about my research, please send me an e-mail: fiona.tamminga@gmail.com."

2.4.2. Types of questions

The attitudinal questions that are included in this questionnaire are mostly answered via a Likert scale. The category of "*attitudinal questions"* is "used to find out what people think, covering attitudes, opinion, beliefs, interests, and values" (Dörnyei 2007:102). "Whether we take […] attitude […] as an entity innate or learned, it is in either case not an inflexible and rigid element in personality (if, in fact, any such elements exist), but rather a certain *range within which responses move*" (Likert 1932:8). The Likert scale or "the method of summated ratings" has been popular amongst applied linguists (Low 1999:503). It is a well-known type of closed-ended questions; they contain a "characteristic statement" and the participants are requested to indicate the level to which they "'agree' or 'disagree'" by marking the accurate response (Dörnyei 2007:104). Each option will be given a number in order to score the answers and these will thereafter be used for the statistic calculations (Dörnyei 2007:105). Originally, Likert (1932) intended the scales to represent only one element such as "approval" and to connect to the statements in a substantial manner (Low 1999:503). However, it appears that the scales are often "made bipolar" with for example "'approval' versus 'disapproval'" and the "interval boundary points" are labelled with indications such as "*I strongly approve*" even though Likert (1932) addresses that this is not necessary (Low 1999:504).

In this survey, the usual bipolar scales of approval and disapproval are substituted by the names of the two people from the story. York will be on the left hand side of the scale, and Rome will be on the right hand side. The scale will have 5 interval boundary points with the middle point being the neutral option. The participants will then have to choose whether the particular idiom in question fits more with York's or with Rome's character or if York or Rome is more likely to display the characteristics described in said idiom. Afterwards, the scores of 1 to 5 for every single idiom will be added and divided by the number of answers, thereby providing an average answer for each idiom with which the answers can be compared.

Furthermore, at the end of the questionnaire, "*factual questions* which are used to find out certain facts out the respondents, such as demographic characteristics […], residential location, marital and socio-economic status, level of education, language learning history, amount of time spent in an L2 environment, etc." will be asked (Dörnyei 2007:102). Dörnyei and Taguchi (2010:16) categorize these questions among the "sensitive" topics; these "facts of life" may bear a social or emotional weight for the participants, so much so that they will succumb to answering questions according to the ""social desirability" bias" that will be discussed in the following section. For this reason, Dörnyei and Taguchi (2010:16) advise to refrain from asking "sensitive questions" unless the project demands it. Schleef (2014:49) recommends placing them at the end of the questionnaire, to avoid discouraging people from finishing it. Another method to dilute the sensitiveness is making the survey anonymous (Dörney and Taguchi 2010:16). Following these recommendations, both tactics are applied in this questionnaire.

2.4.3. Limitations

Even though a questionnaire might seem to be the perfect tool to investigate people's attitudes about the idioms included in this study, Dörnyei and Taguchi (2010:7) caution that "questionnaires have some serious limitations"; therefore, both the limitations and the measures that have been taken to ensure the best possible results will be addressed in the following paragraphs.

First, the "[s]implicity and [s]uperficiality" of the answers is a concern (Dörnyei and Taguchi 2010:7). Participants will fill in their answers independently; hence, the questions need to be self-explanatory and simple enough that everybody can understand them (Dörnyei and Taguchi 2010:7). Furthermore, respondents are usually not inclined to spend much time answering the questions (Dörnyei and Taguchi 2010:7). For these reasons, the questionnaire for this study has been designed in such a way that it clearly leads the participants through the sections of the survey, thereby indicating how far they have come and how much is left to accomplish. This was done by adding *1/3*, *2/3*, and *3/3* to indicate the different idiom question sections, and by adding encouraging words such as *almost done* to the title heads. In addition, the feature that indicates the advancement of the questionnaire in Google Forms was activated. Together with the usage of everyday language and avoidance of any academic level vocabulary or linguistic concepts, these measures will hopefully make sure that the respondents understand the survey.

However, it does not matter how straightforward and easily understandable the survey is, if the participants lack the motivation to answer correctly. Dörnyei and Taguchi (2010:7) warn for "[u]nreliable and [u]nmotivated [r]espondents". Many will not be careful or observant answering the questions, because this is something that most people do not particularly enjoy or receive any benefit from (Dörnyei and Taguchi 2010:7). In their empirical study, Low (1999:528) found that the students taking part in the study would often misinterpret the questions or rather interpret them in their own way by for instance devising a different scale which then acts as the main scale in the survey. In order to keep the lack of attentiveness to a minimum, attempts were made to keep the questionnaire as interesting as possible. The story that was devised as background to the idioms is a typical love story that everybody knows from the cinema. It is straightforward and the premise will be familiar to the participants. Hopefully, the participants will be motivated by the bar indicating their progress as mentioned in the previous paragraph. Furthermore, the layout of Google Forms is modern and colourful. This will also give the appearance of a more fun activity than filling in a business-like survey.

Lastly, Dörnyei and Taguchi (2010:8) mention the "[s]ocial [d]esirability (or [p]restige) [b]ias". The respondents will not always provide answers that are completely true; the results show what the participants "*report* to feel or believe, rather than what they *actually* feel or believe" (Dörnyei and Taguchi 2010:8). This is most probably a result of the "*social desirability* or *prestige bias*, given that the people will know or guess the desired answer for particular questions and they will be tempted to present themselves in the best way possible (Dörnyei and Taguchi 2010:8). In light of this information, the goal of the questionnaire will be hidden from the participants. This means that there will be no mention of sexism or gender bias. They will of course have the opportunity to request more information after completing the survey. If the respondents stay in the dark about the purpose of the research, they will fill in the Likert scales intuitively and not have male or female bias in the back of their minds. To further disguise the purpose of the questionnaire, dummy questions will be added to ensure that the participants will not figure it out and bias their own answers.

2.4.4. Piloting

Dörnyei and Taguchi (2010:54-55) emphasise the importance of feedback during the construction time of the questionnaire. They (2010:54-55) envision an initial piloting and a final piloting. The initial piloting should consist of a small group of people who are willing to help you with this process (Dörnyei and Taguchi 2010:55). They must not all be familiar with the field or with questionnaires in general (Dörnyei and Taguchi 2010:55). They are asked to provide feedback about the survey ranging from general comments to specific wording suggestions (Dörnyei and Taguchi 2010:55). The next stage is the final piloting or "Dress

Rehearsal" as Dörnyei and Taguchi (2010:55) call it. Here, the survey is put in front of a group of participants who are part of the target group (Dörnyei and Taguchi 2010:56).

These stages of piloting were also implemented for this questionnaire; however, I also sent the construction of the whole survey, including the particular idioms that I wanted to use, to my supervisor before distributing the survey. Univ.-Prof. Dr. Keizer gave several helpful suggestions to improve the general questionnaire. Afterwards, the initial pilot was sent to friends and colleagues from the university: two from the linguistic department, one from the economics, and one from the physics department. I had been discussing ideas for my thesis with all of them; thus, they were aware of the purpose of my thesis. For this reason, they were instructed to focus on the general ease of going through the slides, and on the wording of the instructions, stories, and questions. Finally, I asked several friends that I had not discussed my project with to fill in the survey. Their experience was discussed afterwards. The feedback that I received has been valuable information for the improvement of the questionnaire.

3. Results: Questionnaire

3.1. Statistics

The questionnaire was devised in such a way that the results could easily be converted into statistically analysable data. Dörnyei (2007:104) confirms that questionnaires that ask for "very specific pieces of information" are well-suited for "quantitative, statistical analysis". However, Egbert, Larson and Biber (2020:42) emphasize that it is wise to avoid "an overreliance on the results of statistical tests". According to them (2020:40), "we should always strive to choose *minimally sufficient* statistical methods". This is why I will not rely only on presenting the statistical calculations and the resulting numbers. Instead, I will also present a clear overview of the data by showing graphs so as not to "abstract away from the language" too much (Egbert et al. 2020:40). Statistics, even though they are useful, can distract from the research goal and we should be careful to "not use sight of [our] role as linguists" (Egbert et al. 2020:53).

Furthermore, as a consequence of SPSS being so easy to use even for people who do not usually work with statistics, it may be hard to not get lost in the various options that the programme offers. It is sometimes enticing to just try and do the tests without understanding what the underlying principles are. For this reason, this chapter is constructed with a clear structure; it follows the steps that were taken in the specific order that was used to get the results. Moreover, each step is accompanied by an explanation of the statistics that were used. Therefore, it will be easy to follow and understand the process.

3.2. Preparing the data

SPSS software was developed for research in the area of social sciences Larson-Hall (2016:3). Indeed, the official description on the product's website shows that: "[t]he IBM® SPSS® software platform offers advanced statistical analysis, a vast library of machine learning algorithms, text analysis, open source extensibility, integration with big data and seamless deployment into applications" (*IBM* 2022). It was, and still is, according to Larson-Hall (2016:3), the most popular statistical programme in that field at the moment. Furthermore, many researchers with a focus on second language research use SPSS (Larson-Hall 2016:3). The programme has a "graphical user interface (GUI)" with "drop-down menus" that are similar to the ones that programmes such as Microsoft Word offers (Larson Hall 2016:3).

Several steps need to be taken to transform the questionnaire from Google Forms into data that SPSS can work with. First of all, Google Forms does not produce an automatic Excel spreadsheet. Instead this is done through importing the Google spreadsheet into Excel (*Popan*

2017). Second, the data in this file must be adapted so that it is compatible with the SPSS software. This means for instance clear and concise variable names without question marks (*Popan* 2017). Third, spelling should be checked (*Popan* 2017). It makes sense to correct this and make sure that all the names are typed in a similar way. For instance, The Netherlands, the netherlands, and Nederland will have to be corrected to the same country indicator, namely NL in this case. Otherwise, these names will count as different values.

There are different categories of variables and the use of terminology mirrors that of idioms in that there is no consensus: Chetty (*2015*) calls them nominal, ordinal, and scale, while Eddington (2015:7) refers to them as categorical, ordinal, and continuous. Nominal, or categorical, variables do not have an intrinsic ranking (Chetty *2015*); examples from this dataset are nationality, country of origin, and gender. Eddington (2015:7) warns that the number of items that are categorical are countable. However, the category should not be regarded as a number; therefore, it cannot be used for mathematical formulas (Eddington 2015:7). The second category, that of ordinal variables, does have such a ranking (Chetty *2015*). These variables have an arbitrary zero point (Oakes 1998:9). Here, the distance between the variables is not important; it is only the order that must be noted (Eddington 2015:7). Therefore, the Likert scale answers of this questionnaire will receive the degrees of 1 = York, 2 = more York than Rome, 3 = neutral, 4 = more Rome than York, and lastly 5 = Rome. Additionally, the levels of English are rated according to the European Framework of Reference: A1 to C2. 1 = A1, 2 = A2, 3 = B1, 4 = B2, 5 = C1, 6 = C2, and finally 7 = Native speaker. The variables from the third and final category of scale, or continuous, communicate a "meaningful metric" (*Chetty* 2015) in which the distance that can be found between the numbers is equal (Eddington 2015:7). These variables, such as age or years, have a meaningful zero point (Oakes 1998:9).

3.3. P-value

When statistical tests are used in research, the corresponding hypotheses are declared in the "null form" (Oakes 1998:9). This null hypothesis articulates that the data and the population from which this was taken show no difference (Oakes 1998:9). Generally, the goal of the researcher is of course to disprove this and to offer statistical data to show that they do differ (Oakes 1998:9). However, Field (2009:27) cautions that the original hypothesis is not automatically proven, if the null hypothesis is rejected; this only supports it. Indeed, it can generally only be claimed that a certain hypothesis is supported by a study or that an idea is made plausible through evidence (Larson-Hall 2016:59). In order to accomplish this, the null hypothesis may only be rejected if there are five or fewer occurrences out of a 100 that would obtain this result (Oakes 1998:9). This means that the "probability of obtaining these results is less than 0.05" (Oakes 1998:9). In this case, researchers "may claim statistical significance" for their data (Oakes 1998:9). In other words, if the chance of something happening accidentally is 5%, or there is a probability of 0.05, the finding is "[statistically significant]" (Field 2009:50). In that case, the null hypothesis is rejected and the alternative hypothesis can be accepted (Larson-Hall 2016:59).

3.4. Descriptive statistics

A spreadsheet full of data is difficult to interpret. This is where descriptive statistics can be helpful; it allows us to create a general overview of the data (Eddington 2015:9). A score on its own does not carry much meaning, but can become meaningful if it is put into context and compared to the scores that the rest of the population received (Eddington 2015:9) This "average test score" is called the mean and "is a measure of central tendency" (Eddington 2015:9). The "central tendency" denotes the "centre of a frequency distribution" (Field 2009:20). Another "measure of central tendency" is the median: "the middle score" (Eddington 2015:10). If the scores are put into order from the lowest to the highest one, the median is the score in the middle (Eddington 2015:10). The final "measure of central tendency" is the mode: the score that most often occurs in the dataset (Eddington 2015:11).

3.4.1. Participants

It was important to get enough native and non-native speakers of English to participate in this study to be able to create groups to compare statistically. First of all, the link of the questionnaire was sent to friends, family, and university colleagues. They were asked if they could help with my research and fill in the questionnaire. Furthermore, sharing the link of the questionnaire was encouraged. This way the link to the survey found its way to a WhatsApp group of international chess players, a group of American friends, and groups of work colleagues amongst others. Second of all, the questionnaire was posted in various Facebook groups for example "Nederlanders in Wenen", a group aimed at Dutch expats in Vienna with 1,743 members, "Americans in Vienna", a group created for American expats in Vienna with 2,445 members, and "Anglistik and Amerikanistik" for both the University of Vienna, 3,998 members, and the University of Graz, 3,014 members. Finally, potential participants on several pages with a focus on linguistics were targeted, because it was estimated that they would be more likely to help a fellow linguist than people that did not share this interest. The largest pages were "Linguistics" with 47,006 members and "The Language Nerds" with 403,905 members.

These numbers appear to be large; however, this does not mean that these people are all active members. Moreover, the Facebook algorithm will randomly show or hide posts. Therefore, I actively commented on people that reacted to my post saying they had done the questionnaire or that asked questions about the process. This way, I hoped to influence the popularity of the post. Moreover, my friends with Facebook tried to click on the like button of the post to increase the visibility as well.

3.4.2. Number of participants

All in all, 239 people filled in the questionnaire online. The specifics can be seen in the following Table 8.

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	160	66,9	66,9	66,9
	Male	69	28,9	28,9	95,8
	Non-binary	5	2,1	2,1	97,9
	Prefer not to say	5	2,1	2,1	100,0
	Total	239	100,0	100,0	

Table 8: Original number of participants

Of the 239 participants in total, 160 people identified as female, 69 as male, and five as non-binary. Furthermore, five participants preferred not to publish their gender. Because I would like to compare the female group and the male group against each other, I will not take the non-binary and the group that preferred to remain unidentified into account for this study. It would of course have been possible to compare the non-binary group to the other results; however, five people is not enough to base the calculations on. A large set of non-binary participants may offer further insight and could be a prospective research project. As a result, the final group of participants counts 229 participants as Table 9 shows. 160 or 69.9% of these participants identify as female, whereas 60 or 30.1 % of them identify as male.

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	160	69,9	69,9	69,9
	Male	69	30,1	30,1	100,0
	Total	229	100,0	100,0	

Table 9: Actual particpants

3.4.3. Age

It appears that the age of the participants has a wide range based on Table 10 and Figure 3. There are 227 valid answers; two participants did not share their age. The mean age is 38.48

with a standard deviation of 14.00. The youngest participant of the questionnaire is 16, while the oldest is 82. Furthermore, the mode, or the age that is most frequent among the participants, is 27. Overall, there is a higher concentration of participants in their late twenties to early thirties which possibly occurred due to friends and university colleagues being more likely to answer a questionnaire for someone they know than random strangers of the internet.

Statistics

Age

N	Valid	227
	Missing	2
Mean		38,48
Median		35,00
Mode		27
Std. Deviation		14,002
Minimum		16
Maximum		82

Table 10: Age of the participants

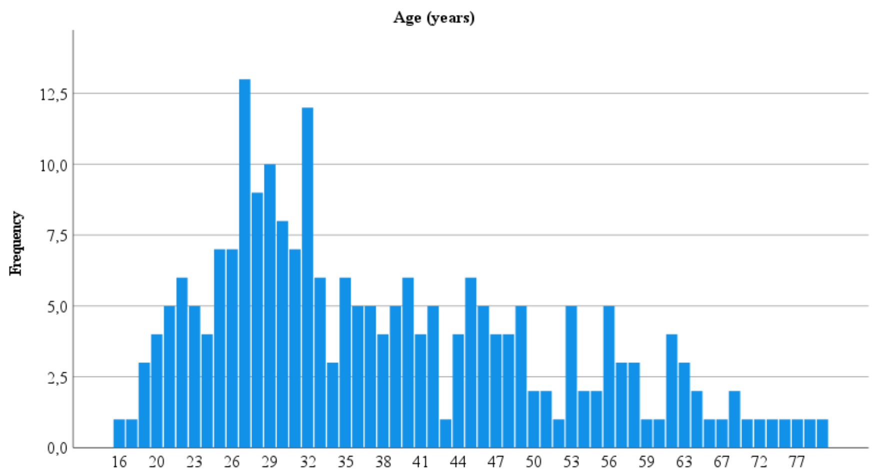

Figure 3: Age of the participants

3.4.4. Nationality

There are 42 nationalities among the participants. These are only depicted in the following Figure 4 when they account for more than 1% of the answers. The rest is collected under the bar called "Other". This already gives an impression of the most frequent native languages with the USA, 63, Austria, 37, the UK, 12, and the Netherlands, 39, belonging to the most frequent countries of origin.

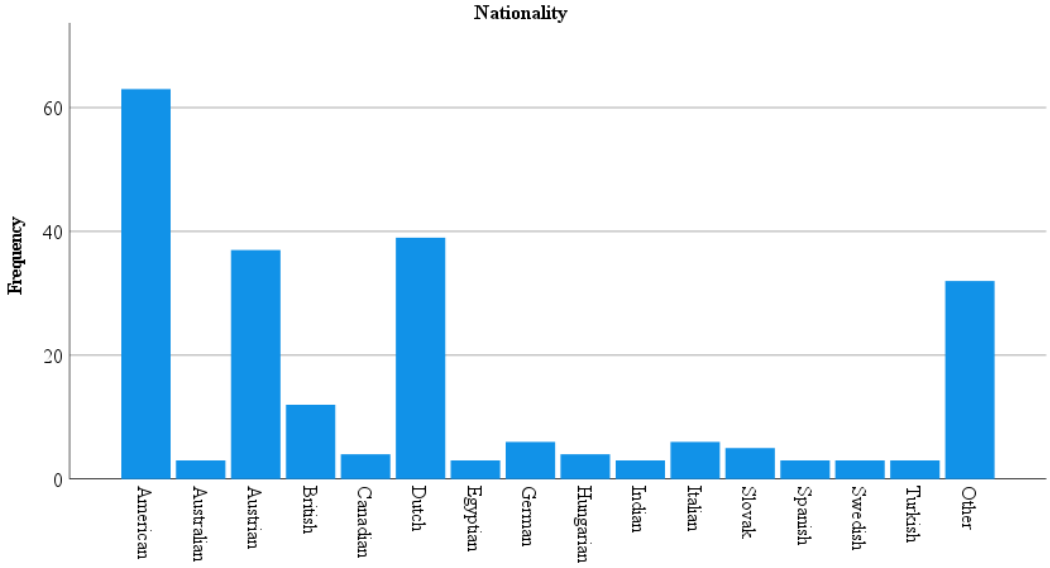

Figure 4: Nationality of the participants

3.4.5. Level of English

The female group consists of 66 native speakers and 94 non-native speakers of English. This makes up 28.82% and 41.05% respectively of the total number of participants, which can be seen in Figure 5 below. On the contrary, the male group does not boast as many members with 24 native speakers and 45 non-native speakers. These two groups account for 10.48% and 19.65% of the total number of participants respectively. It follows that the group of participants can also be divided in native speakers, 90 members or 39.30% of the total number, and non-native speakers, 139 members or 60.7%.

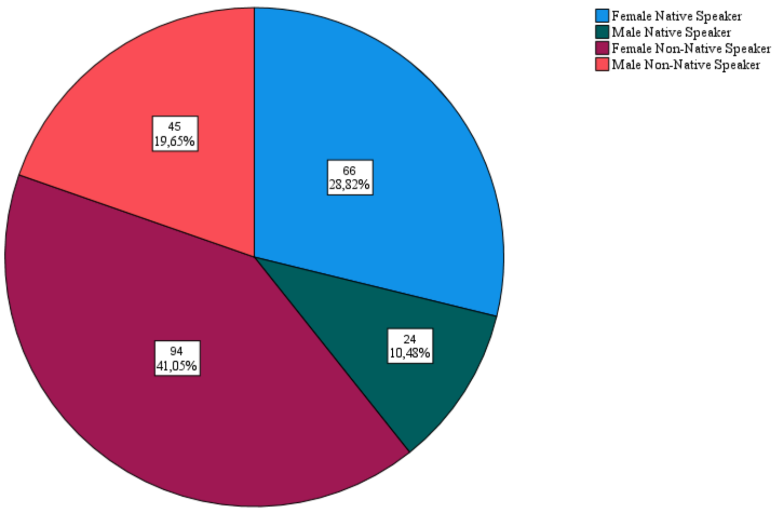

Figure 5: Group composition

As mentioned, the group of non-native speakers consists of 139 people. They were asked to estimate their own level of English according to the European Framework of Reference: A1 starter, A2 elementary, B1 intermediate, B2 upper intermediate, C1 expert, and C2 mastery. The results can be seen in the two Tables 11 and 12 below. On average, the level of English is almost C1, with a mean of 4.79 and a standard deviation of 1.23. 91.3% of the participants judge themselves to have a B1 level or higher, while 66.1% of them even have a C1 or higher.

English Level

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A1	3	,0	2,2	2,2
	A2	6	,0	4,3	6,5
	B1	9	,0	6,5	13,0
	B2	29	,0	20,9	33,9
	C1	44	,0	31,7	65,6
	C2	48	,0	34,4	100,0
	Total	139	,0	100,0	

Table 11: English level of non-native speakers

Statistics

English Level

N	Valid	139
	Missing	0
Mean		4,7914
Median		5,0000
Mode		6,00
Std. Deviation		1,23050
Minimum		1,00
Maximum		6,00

Table 12: English level of non-native speakers statistics

3.5. Normality

It is crucial to discover whether the data is normally distributed or not, because various statistical tests have a normal distribution as a requirement (Eddington 2015:16), while other tests must be used with data that does not fit this description. Every data point is located at a particular distance from the mean; the average distance these data points are away from the mean represent the standard deviation (Eddington 2015:15). In a perfect normal distribution, 68% of the scores would fall within the first standard deviation, divided in a section both above and below the mean (Eddington 2015:15). 95% of the scores would fall within the second standard deviation (Eddington 2015:15). The value of the standard deviation indicates how disperse these data points are (Eddington 2015:15). An imperfect distribution can be caused due to "(1) a lack of symmetry (called skew) and (2) pointiness (called kurtosis)" (Field 2009:19). These two values would be 0 in a normal distribution (Field 2009:19). Through Shapiro-Wilk and Kolmogorov-Smirnov tests the distribution of the research data can be examined and compared with that of a normal distribution; this shows whether there is a significant difference between them (Eddington 2015:18; *Grande* 2015). The results from the idioms are normally distributed when the p-value is 0.05 or higher. However, as can be seen in Table 13, the data is not "statistically significantly different from a normal distribution" for these variables (*Grande* 2015), because the significance is <0.001 and thus lower than the p-value in all instances.

Tests of Normality

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
wear the trousers	,291	229	<,001	,789	229	<,001
take under wing	,182	229	<,001	,865	229	<,001
in drivers seat	,292	229	<,001	,776	229	<,001
pick up tab	,333	229	<,001	,757	229	<,001
call the shots	,199	229	<,001	,866	229	<,001
stick neck out	,189	229	<,001	,907	229	<,001
final say	,203	229	<,001	,889	229	<,001
put head on block	,255	229	<,001	,884	229	<,001
pillar of strength	,205	229	<,001	,891	229	<,001
run the show	,206	229	<,001	,878	229	<,001
out of their league	,261	229	<,001	,873	229	<,001
breaking their back	,186	229	<,001	,906	229	<,001
being on the ball	,207	229	<,001	,886	229	<,001
cooking up a storm	,264	229	<,001	,881	229	<,001
smart as a whip	,270	229	<,001	,865	229	<,001
pulling their weight	,256	229	<,001	,886	229	<,001
beck and call	,200	229	<,001	,885	229	<,001
quick on the trigger	,262	229	<,001	,858	229	<,001
holding their tongue	,224	229	<,001	,872	229	<,001
being up to snuff	,340	229	<,001	,799	229	<,001
wandering eye	,265	229	<,001	,869	229	<,001
love em and leave em	,317	229	<,001	,817	229	<,001
apple of their partners eye	,173	229	<,001	,871	229	<,001
the better half	,271	229	<,001	,819	229	<,001
trophy person	,198	229	<,001	,880	229	<,001
of easy virtue	,328	229	<,001	,792	229	<,001
says the old ball and chain	,232	229	<,001	,878	229	<,001
tied to the apron strings	,191	229	<,001	,891	229	<,001
got hitched	,323	229	<,001	,816	229	<,001
says a match made in heaven	,199	229	<,001	,896	229	<,001

Table 13: Tests of normality

3.6. Reliability

There are variables in the hypothesis of this study that cannot be measured directly; these are called latent variables (*DATAtab* 2021). The six latent variables of this study are the six groups of the Ambivalent Sexism Theory: paternalism, gender differentiation, and heterosexual, with both a hostile and benevolent variant for each of those. The so-called scale, which is a group of questions, is used in order to make it possible to measure these variables (*DATAtab* 2021). This scale is then used to measure the latent variable (*DATAtab* 2021). Thus, the six scales of this research consist of the five questions each of them contains. The responses to this group of questions should be "highly correlated" (*Datatab* 2021). If this is indeed the case, the scale has a "high internal consistency" (*Datatab* 2021). This can be tested by using Chronbach's alpha; this is a measure of the reliability or "internal consistency (homogeneity)" of Likert scale questionnaires with multiple questions (*Glen* 2022). A score between 0.6 and 0.7 is questionable, while everything above 0.7 is deemed acceptable (*Glen* 2022). A score below 0.5 however is found unacceptable (*Glen* 2022). It follows that, generally, the higher the better when it comes to this score (*Glen* 2022).

Reliability Statistics

Hostile Paternalism

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,657	,661	5

Benevolent Paternalism

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,992	1,000	5

Hostile Gender Differentiation

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,998	1,000	5

Benevolent Gender Differentiation

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,998	1,000	5

Hostile Heterosexuality

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,997	1,000	5

Benevolent Heterosexuality

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,999	1,000	5

Table 14: Chronbach's Alpha for all scales

In Table 14, it can be seen that all scales except one have an excellent internal consistency. They all achieved Chronbach's alphas of at least 0.992. The only exception to this is the scale of hostile paternalism which has a questionable correlation with a Chronbach's alpha of 0.657.

Reliability Statistics

Paternalism

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,579	,607	10

Gender Differentiation

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,998	1,000	10

Heterosexuality

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,998	1,000	10

Table 15: Chronbach's Alpha for the three concepts

This pattern is repeated in Table 15 where the hostile and benevolent concepts were tested together. Here, gender differentiation and heterosexuality again achieved excellent scores with a Chronbach's alpha of 0.998 for both. Paternalism, as expected, has the lowest score. The low score of the scale of hostile paternalism influences the complete concept. Thus, the overall concept only receives a poor Chronbach's alpha of 0.579. However, this is still not unaccepTable according to the scale. Therefore, overall, the reliability of these questions is sufficient.

3.7. Inferential statistics

In statistics, parametric and non-parametric tests are used on data. Parametric tests work with the assumption that the variables are ratio-scaled or interval-scaled (Oakes 1998:9). Furthermore, it is often assumed that the data should have a normal distribution; however, parametric tests can be done with different types of distribution (Oakes 1998:9). Because the data of this research project consists of ordinal, "rank-ordered scales" variables (Oakes 1998:9), non-parametric tests will be used. According to Oates (1998:9) these tests work well for this type of data, because they do not rely on normal distribution of a population and they work well even with small sample sizes. Most of the non-parametric tests function on the basis of "ranking" the data; in other words, the system finds the lowest score and accords it with the rank of 1, then they accord the next score with a rank of 2 (Field 2009:540). This process continues until all the higher scores are defined by higher ranks, and all the lower scores by lower ranks (Field 2009:540). Thereafter, the analysis is not based on the actual data that was put into the system, but on the calculated ranks (Field 2009:540).

3.7.1. Independent t-test

The Mann-Whitney test is the "nonparametric replacement for an independent t-test" and is suitable for ordinal data (Eddington 2015:54). It is used when a comparison is needed between two independent groups (Oates 1998:16-17). This test creates a comparison of "ordinal rating scales" between two groups based on their scoring "above and below the median" (Oakes 1998:17). The observations that are the input must not be dependent on each other; in other words, each participant should only score once in a group, otherwise their two scores would have a correlation (Eddington 2015:54). Furthermore, each participant should only belong to one single group, otherwise their scoring in the one group would not be independent of their scoring in the other group (Eddington 2015:54).

To discover whether gender is a deciding factor in the answers to the questions of the survey, a Mann-Whitney test was performed on the scoring of the idioms. One independent group consisted of only female and the other of only male participants. The results are displayed in the following Table 16. It appears that only in six out of 30 instances, or 20%, the answers of the female group are significantly different from those of the male group. Of the ten idioms that belong to paternalism, four, or 40%, show a difference between the female and male answers. Furthermore, the ten idioms that describe gender differentiation, two, or 20%, indicate a difference, while heterosexuality does not show any differences between the answers of the two groups.

Hypothesis Test Summary Category Gender

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The distribution of wear the trousers is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,139	Retain the null hypothesis.
2	**The distribution of final say is the same across categories of Gender.**	**Independent-Samples Mann-Whitney U Test**	**,045**	**Reject the null hypothesis.**
3	**The distribution of in drivers seat is the same across categories of Gender.**	**Independent-Samples Mann-Whitney U Test**	**,050**	**Reject the null hypothesis.**
4	The distribution of run the show is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,992	Retain the null hypothesis.
5	The distribution of call the shots is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,789	Retain the null hypothesis.
6	**The distribution of take under wing is the same across categories of Gender.**	**Independent-Samples Mann-Whitney U Test**	**,048**	**Reject the null hypothesis.**

7	The distribution of stick neck out is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,762	Retain the null hypothesis.
8	The distribution of pick up tab is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,807	Retain the null hypothesis.
9	**The distribution of put head on block is the same across categories of Gender.**	**Independent-Samples Mann-Whitney U Test**	**,008**	**Reject the null hypothesis.**
10	The distribution of pillar of strength is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,075	Retain the null hypothesis.
11	The distribution of out of their league is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,709	Retain the null hypothesis.
12	The distribution of quick on the trigger is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,744	Retain the null hypothesis.
13	**The distribution of being up to snuff is the same across categories of Gender.**	**Independent-Samples Mann-Whitney U Test**	**,010**	**Reject the null hypothesis.**
14	The distribution of being on the ball is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,779	Retain the null hypothesis.
15	The distribution of smart as a whip is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,719	Retain the null hypothesis.
16	The distribution of breaking their back is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,242	Retain the null hypothesis.
17	The distribution of pulling their weight is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,268	Retain the null hypothesis.
18	The distribution of cooking up a storm is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,454	Retain the null hypothesis.
19	**The distribution of beck and call is the same across categories of Gender.**	**Independent-Samples Mann-Whitney U Test**	**,039**	**Reject the null hypothesis.**
20	The distribution of holding their tongue is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,109	Retain the null hypothesis.

21	The distribution of wandering eye is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,945	Retain the null hypothesis.
22	The distribution of love em and leave em is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,361	Retain the null hypothesis.
23	The distribution of of easy virtue is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,959	Retain the null hypothesis.
24	The distribution of tied to the apron strings is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,107	Retain the null hypothesis.
25	The distribution of says the old ball and chain is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,218	Retain the null hypothesis.
26	The distribution of apple of their partners eye is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,391	Retain the null hypothesis.
27	The distribution of the better half is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,575	Retain the null hypothesis.
28	The distribution of trophy person is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,180	Retain the null hypothesis.
29	The distribution of got hitched is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,901	Retain the null hypothesis.
30	The distribution of says a match made in heaven is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	,930	Retain the null hypothesis.

Table 16: Results Mann-Whitney U test for gender

Furthermore, to find out whether being a native speaker influences the answers to the questions of the survey, another Mann-Whitney test was performed on the scoring of the idioms. One independent group consisted of only native and the other of only non-native speakers. The results are displayed in the following Table 17. It appears that again only in six out of 30 instances, or 20%, the answers of the L1 group are significantly different than those of the L2 group. Of the ten idioms that belong to paternalism, three, or 30%, show a difference between the two groups. Furthermore, of the ten idioms that describe gender differentiation, two, or

20%, indicate a difference, while heterosexuality only shows one, or 10%, idiom with a significant difference.

Hypothesis Test Summary Category English Native

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The distribution of wear the trousers is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,318	Retain the null hypothesis.
2	The distribution of final say is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,981	Retain the null hypothesis.
3	The distribution of in drivers seat is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,848	Retain the null hypothesis.
4	The distribution of run the show is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,284	Retain the null hypothesis.
5	**The distribution of call the shots is the same across categories of English Native.**	**Independent-Samples Mann-Whitney U Test**	**,008**	**Reject the null hypothesis.**
6	**The distribution of take under wing is the same across categories of English Native.**	**Independent-Samples Mann-Whitney U Test**	**,007**	**Reject the null hypothesis.**
7	The distribution of stick neck out is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,934	Retain the null hypothesis.
8	**The distribution of pick up tab is the same across categories of English Native.**	**Independent-Samples Mann-Whitney U Test**	**<,001**	**Reject the null hypothesis.**
9	The distribution of put head on block is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,437	Retain the null hypothesis.
10	The distribution of pillar of strength is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,163	Retain the null hypothesis.
11	The distribution of out of their league is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,187	Retain the null hypothesis.

12	The distribution of quick on the trigger is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,153	Retain the null hypothesis.
13	The distribution of being up to snuff is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,229	Retain the null hypothesis.
14	The distribution of being on the ball is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,226	Retain the null hypothesis.
15	The distribution of smart as a whip is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,894	Retain the null hypothesis.
16	The distribution of breaking their back is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,295	Retain the null hypothesis.
17	The distribution of pulling their weight is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,949	Retain the null hypothesis.
18	**The distribution of cooking up a storm is the same across categories of English Native.**	**Independent-Samples Mann-Whitney U Test**	**,033**	**Reject the null hypothesis.**
19	The distribution of beck and call is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,824	Retain the null hypothesis.
20	**The distribution of holding their tongue is the same across categories of English Native.**	**Independent-Samples Mann-Whitney U Test**	**,019**	**Reject the null hypothesis.**
21	The distribution of wandering eye is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,160	Retain the null hypothesis.
22	The distribution of love em and leave em is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,672	Retain the null hypothesis.
23	The distribution of of easy virtue is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,876	Retain the null hypothesis.

24	The distribution of tied to the apron strings is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,152	Retain the null hypothesis.
25	The distribution of says the old ball and chain is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,058	Retain the null hypothesis.
26	The distribution of apple of their partners eye is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,224	Retain the null hypothesis.
27	**The distribution of the better half is the same across categories of English Native.**	**Independent-Samples Mann-Whitney U Test**	**,011**	**Reject the null hypothesis.**
28	The distribution of trophy person is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,276	Retain the null hypothesis.
29	The distribution of got hitched is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,199	Retain the null hypothesis.
30	The distribution of says a match made in heaven is the same across categories of English Native.	Independent-Samples Mann-Whitney U Test	,412	Retain the null hypothesis.

Table 17: Results Mann-Whitney U Test for native and non-native speakers

Based on these results, the research questions that ask whether the results from the participants that identify as female are significantly different from those that identify as male and whether native speakers are more aware of this hidden bias compared to non-native speakers can already be answered in a preliminary manner. As the answers to the questionnaire only differ significantly in 20% of the idioms for both the test comparing the male and the female group as well as the one comparing the native and the non-native speaker group, it is likely that the research question will be answered negatively. 80% of the idioms show no significant difference between the answers of these groups; therefore, further tests will probably confirm this. Moreover, if the answers between the L1 and the L2 group continue to show no differences, the amount of awareness of a hidden gender bias in idioms will likely be similar.

3.7.2. Paired-samples t-test

In contrast to the independent t-test, the paired-samples t-test is used when the groups are dependent on each other (Larson-Hall 2010:242). This could be because the group was measured at two different points in time, or, which is applicable to this data, because the group was tested on two related issues (Larson-Hall 2010:242). In both cases, the difference in performance can be investigated (Larson-Hall 2010:242). Consequently, the same people will appear twice in the results; therefore, it must be assumed that these groups are dependent (Larson-Hall 2010:242). The non-parametric paired t-test is called the Wilcoxon signed-rank test (Eddington 2015:60).

The Likert scale of the survey indicated the male bias with the number 1, whereas the female bias was signified by the number 5. It is thus safe to say that a lower scoring of a participant will indicate a personal male bias; the lower the score, the higher the bias. Remember, the story on which the participants were asked to base their answers to the idioms was linguistically completely gender neutral. York was given traditionally masculine stereotypical character traits and behaviours, while Rome was given the feminine equivalent. Thus, someone answering that York must wear the trousers acted out of a male bias. Therefore, if it can be shown that the participants scored the hostile elements of the type of sexism lower than their benevolent parts, this would indicate that hostile sexism is more recognized than benevolent sexism. For this reason, the scores of the six different types of sexism were added to each other. This way, each participant has one number representing one type of sexism. Thereafter, the score of each of the three pairs of hostile and benevolent sexism was compared for the 229 participants with a Wilcoxon signed-rank test. The results can be found in the Table 18 below.

Hypothesis Test Summary all Participants

Hostile and Benevolent Paternalism

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The median of differences between SUM PAT HOS and SUM PAT BEN equals 0.	Related-Samples Wilcoxon Signed Rank Test	,000	Reject the null hypothesis.

Hostile and Benevolent Gender Differentiation

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The median of differences between SUM GD DIFF HOS and SUM GD DIFF BEN equals 0.	Related-Samples Wilcoxon Signed Rank Test	,000	Reject the null hypothesis.

Hostile and Benevolent Heterosexuality

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The median of differences between HETSEX HOS and HETSEX BEN equals 0.	Related-Samples Wilcoxon Signed Rank Test	,000	Reject the null hypothesis.

Table 18: Results Wilcoxon signed-rank test for the hostile and benevolent type

All three tests show a significant difference with a p-value of 0.00; this means that the null hypothesis, namely that the two groups that were tested do not show any differences, can be rejected. For paternalism, 154 pairs of the 229, or 67%, show a higher score for the benevolent compared to the hostile variant, 43 pairs, or 19%, show a lower score and 32, or 14%, scores were tied. Furthermore, for gender differentiation, 167 pairs of the 229, or 73%, show a higher score, 27, or 12 %, a lower score, and 36, or 15%, a tie. Finally, for heterosexuality, 156 pairs of the 229 or 68% show a higher score, 36, or 15%, a lower score, and 38, or 17% a tie. To provide a structured overview, these results are summarized in Table 19 below.

Whole Group		**PAT**			**GD**			**HET**	
	Higher	**Tie**	**Lower**	**Higher**	**Tie**	**Lower**	**Higher**	**Tie**	**Lower**
Number	154	32	43	167	36	27	156	38	36
%	67%	14%	19%	73%	15%	12%	68%	17%	15%

Table 19: Results Wilcoxon signed-ranks test for the whole group

In view of these results, it can be said that there is a difference between hostile and benevolent sexism. The significant difference that appears in the paired samples t-test is clearly presented in Table 19. The percentage of the pairs that score the hostile sexism variant lower than the benevolent variant lies around 70% for the whole group. Indeed, it follows that hostile sexism is recognized more than benevolent sexism. Through the scoring of hostile sexism, the general level of male bias of the participants can be observed. If these people would recognize the benevolent sexism in the idioms as well as the hostile type, they would not score them differently.

Additionally, it is also interesting to see whether the two separate groups of people identifying either as male or as female show the same tendency. For this reason, the same test was done on these groups. The results are displayed in the following Tables 20-22.

Hypothesis Test Summary Female Group

Hostile and Benevolent Paternalism

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The median of differences between SUM PAT HOS and SUM PAT BEN equals 0.	Related-Samples Wilcoxon Signed Rank Test	,000	Reject the null hypothesis.

Hostile and Benevolent Gender Differentiation

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The median of differences between SUM GD DIFF HOS and SUM GD DIFF BEN equals 0.	Related-Samples Wilcoxon Signed Rank Test	,000	Reject the null hypothesis.

Hostile and Benevolent Heterosexuality

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The median of differences between HETSEX HOS and HETSEX BEN equals 0.	Related-Samples Wilcoxon Signed Rank Test	<,001	Reject the null hypothesis.

Table 20: Results Wilcoxon signed-rank test for the female group

Hypothesis Test Summary Male group

Hostile and Benevolent Paternalism

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The median of differences between SUM PAT HOS and SUM PAT BEN equals 0.	Related-Samples Wilcoxon Signed Rank Test	,005	Reject the null hypothesis.

Hostile and Benevolent Gender Differentiation

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The median of differences between SUM GD DIFF HOS and SUM GD DIFF BEN equals 0.	Related-Samples Wilcoxon Signed Rank Test	<,001	Reject the null hypothesis.

Hostile and Benevolent Heterosexuality

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The median of differences between HETSEX HOS and HETSEX BEN equals 0.	Related-Samples Wilcoxon Signed Rank Test	<,001	Reject the null hypothesis.

Table 21: Results Wilcoxon signed-rank test for the male group

Groups		PAT			GD			HET	
	Higher	**Tie**	**Lower**	**Higher**	**Tie**	**Lower**	**Higher**	**Tie**	**Lower**
Female	115	20	25	122	19	20	112	25	24
%	72%	13%	15%	76%	12%	12%	70%	15%	15%
Male	39	12	18	46	17	7	45	13	12
%	57%	17%	26%	66%	24%	10%	65%	18%	17%

Table 22: Results Wilcoxon signed-ranks test for the female and male group

As with the previous tests, there is a significant difference between hostile and benevolent sexism. This means that both the people who identified as male and those who identified as female differentiated between the two types of sexism. The numbers are comparable to the results of the previous paired samples test that ran the test on the complete group of participants. Thus, the combined results of benevolent sexism were scored higher than those of hostile sexism, again portraying a male bias. I would argue that these Tables 20-22 also support the previous independent samples test where it was shown that there is no significant difference in answers between the male and the female group. Indeed, both groups reject the null hypothesis for all three categories of idioms, thereby showing that their answers are similar.

To finish this part of the analysis, the difference between native and non-native speakers must also be investigated. For this reason, the same test was done on these groups. The results are displayed in the following Tables 23-25.

Hypothesis Test Summary Native group

Hostile and Benevolent Paternalism

	Null Hypothesis	Test	Sig.^{a,b}	Decision
1	The median of differences between SUM PAT HOS and SUM PAT BEN equals 0.	Related-Samples Wilcoxon Signed Rank Test	<,001	Reject the null hypothesis.

Hostile and Benevolent Gender Differentiation

	Null Hypothesis	Test	Sig.^{a,b}	Decision
1	The median of differences between SUM GD DIFF HOS and SUM GD DIFF BEN equals 0.	Related-Samples Wilcoxon Signed Rank Test	<,001	Reject the null hypothesis.

Hostile and Benevolent Heterosexuality

	Null Hypothesis	Test	Sig.^{a,b}	Decision
1	The median of differences between HETSEX HOS and HETSEX BEN equals 0.	Related-Samples Wilcoxon Signed Rank Test	<,001	Reject the null hypothesis.

Table 23: Results Wilcoxon signed-rank test for the native group

Hypothesis Test Summary Non-Native group

Hostile and Benevolent Paternalism

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The median of differences between SUM PAT HOS and SUM PAT BEN equals 0.	Related-Samples Wilcoxon Signed Rank Test	<,001	Reject the null hypothesis.

Hostile and Benevolent Gender Differentiation

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The median of differences between SUM GD DIFF HOS and SUM GD DIFF BEN equals 0.	Related-Samples Wilcoxon Signed Rank Test	<,001	Reject the null hypothesis.

Hostile and Benevolent Heterosexuality

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The median of differences between HETSEX HOS and HETSEX BEN equals 0.	Related-Samples Wilcoxon Signed Rank Test	<,001	Reject the null hypothesis.

Table 24: Results Wilcoxon signed-rank test for the non-native group

Group		**PAT**			**GD**			**HET**	
	Higher	**Tie**	**Lower**	**Higher**	**Tie**	**Lower**	**Higher**	**Tie**	**Lower**
Native	64	12	14	62	18	11	65	16	10
%	71%	13%	16%	68%	20%	12%	72%	17%	11%
Non-native	90	20	29	106	18	16	92	22	26
%	66%	14%	10%	76%	13%	11%	66%	16%	17%

Table 25: Results Wilcoxon signed-ranks test for the native and non-native group

As with the previous tests, there is a significant difference between the hostile and benevolent variation of sexism for paternalism, gender differentiation, and heterosexuality. This means that both native and non-native speakers differentiated between these two categories. Similarly to the previous Wilcoxon Signed Ranks tests, approximately 70% of the participants scored the first lower than the latter. In a likewise manner to the female and the male group, I would claim that these results support those of the independent samples test; both indicate that there are no significant differences between the groups' answers. Therefore, the native and the non-native group can be said to have given similar answers.

The results from these paired samples t-tests provide an answer to the research question: is the gender bias more pronounced in the results from idioms containing hostile sexism compared to those containing benevolent sexism. The answers from the group as a whole, as well as from the female and the male participants, and from the native and non-native speakers separately point out that there is an overall male bias when comparing the idioms representing hostile sexism with those representing benevolent sexism. To put it another way, the answers pertaining to hostile sexism were consistently scored lower than those belonging to benevolent sexism. Thus, the participants displayed a more pronounced gender bias when dealing with hostile sexism as opposed to benevolent sexism, thereby answering the research question.

Furthermore, the research question whether L1 speakers are more aware of this hidden bias compared to L2 speakers can be answered as well. These groups do not behave differently from the group as a whole; they also show a larger awareness of hostile sexism compared to benevolent sexism by scoring the first lower than the latter. In conclusion, the whole group, the group divided into people who identify as male or as female, and the group divided into native and non-native speakers all show similar behaviour and thus a similar level of awareness.

3.7.3. ANOVA

The previous paragraphs discussed the t-test as a way to discover whether the scoring of one group is significantly different compared to another group; however, this was only possible for two groups. The solution to this problem is the one-way ANOVA. Eddington (2015:65) defines the one-way ANOVA as "a t-test on steroids because it can be used to see if two *or more* groups differ from each other". ANOVA does not only compute the mean of each group, it also examines the "variance of the scores" or the way they are "spread out from the mean" (Eddington 2015:65). This is where ANOVA gets their name: "[analysis of variance]" (Eddington 2015:65). The non-parametric one-way ANOVA test is called the Kruskal-Wallis test (Eddington 2015:69). This test is also suitable for ordinal data (Eddington 2015:69).

In other words, the Kruskal-Wallis test will find out whether the four different groups, namely the female native speakers, the male native speakers, the female non-native speakers, and the male non-native speakers answered the questions of the survey differently. The null-hypothesis here will assume that there is no difference across the groups. In order to reject this, the significance level will have to be 0.05 or lower.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The distribution of wear the trousers is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,330	Retain the null hypothesis.

2	The distribution of final say is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,253	Retain the null hypothesis.
3	The distribution of in drivers seat is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,215	Retain the null hypothesis.
4	The distribution of run the show is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,297	Retain the null hypothesis.
5	The distribution of call the shots is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,068	Retain the null hypothesis.
6	**The distribution of take under wing is the same across categories of GROUP.**	**Independent-Samples Kruskal-Wallis Test**	**,012**	**Reject the null hypothesis.**
7	The distribution of stick neck out is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,943	Retain the null hypothesis.
8	**The distribution of pick up tab is the same across categories of GROUP.**	**Independent-Samples Kruskal-Wallis Test**	**<,001**	**Reject the null hypothesis.**
9	The distribution of put head on block is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,053	Retain the null hypothesis.
10	The distribution of pillar of strength is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,139	Retain the null hypothesis.
11	The distribution of out of their league is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,482	Retain the null hypothesis.
12	The distribution of quick on the trigger is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,309	Retain the null hypothesis.
13	**The distribution of being up to snuff is the same across categories of GROUP.**	**Independent-Samples Kruskal-Wallis Test**	**,020**	**Reject the null hypothesis.**
14	The distribution of being on the ball is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,562	Retain the null hypothesis.
15	The distribution of smart as a whip is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,746	Retain the null hypothesis.

16	The distribution of breaking their back is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,210	Retain the null hypothesis.
17	The distribution of pulling their weight is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,747	Retain the null hypothesis.
18	The distribution of cooking up a storm is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,123	Retain the null hypothesis.
19	The distribution of beck and call is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,094	Retain the null hypothesis.
20	**The distribution of holding their tongue is the same across categories of GROUP.**	**Independent-Samples Kruskal-Wallis Test**	**,037**	**Reject the null hypothesis.**
21	The distribution of wandering eye is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,489	Retain the null hypothesis.
22	The distribution of love em and leave em is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,221	Retain the null hypothesis.
23	The distribution of of easy virtue is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,725	Retain the null hypothesis.
24	The distribution of tied to the apron strings is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,194	Retain the null hypothesis.
25	**The distribution of says the old ball and chain is the same across categories of GROUP.**	**Independent-Samples Kruskal-Wallis Test**	**,044**	**Reject the null hypothesis.**
26	The distribution of apple of their partners eye is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,626	Retain the null hypothesis.
27	The distribution of the better half is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,213	Retain the null hypothesis.
28	The distribution of trophy person is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,353	Retain the null hypothesis.
29	The distribution of got hitched is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,734	Retain the null hypothesis.

30	The distribution of says a match made in heaven is the same across categories of GROUP.	Independent-Samples Kruskal-Wallis Test	,753	Retain the null hypothesis.

Table 26: Results Kruskal-Wallis test

As can be observed in Table 26, only five out of 30 idioms, or 16.7% of the idioms, indicate a significant difference between the four groups of this study. Two questions from the category paternalism, two from gender differentiation, and one from heterosexuality were answered in such a way by the four different groups that they indeed show a significant difference in the results. The other 25 idioms do not indicate such a difference and were therefore answered in a similar manner by all groups.

Now, the research question whether the results from the participants that identify as female significantly different from those that identify as male can be answered. First of all, the results from the non-parametric independent t-test show that female participants answered the questions differently from male participants in only 20% of the questions. Secondly, based on the paired samples t-test, the group identifying as male as well as the group identifying as female answered the questions with a male bias. The two groups of participants show no significant difference in their answers; therefore, I would argue that their results are similar. Finally, the non-parametric one-way ANOVA from this section confirms that there are no significant differences between the answers for the majority of the idioms from the four separate groups: male native speakers, male non-native speakers, female native speakers, and female non-native speakers. In essence, the differences that are found between the two groups are too small to be called overall significant. Moreover, the groups display similarities that are significant in their number. Therefore, the research question must be answered in the negative.

4. Results: Corpus data

The second method of quantitative research of this thesis is based on data collection from the COHA and an Excel template provided by Univ.-Prof. Mag. Dr. Ritt. The template contains several pre-set calculations for the data from the COHA. This will result in several graphs detailing the diachronic development of the data i.e. the occurrences of the chosen keywords from the idioms that are under investigation.

The main function of the COHA that is used for this study is the ability to find "lemmatized and part-of-speech tagged data" (Alatrash, Schlechtweg and Schulte im Walde 2021:3). For verbs, the verb in capitals is used to indicate the lemma followed by an underscore with the letter *v* for verb. Similarly, for nouns, the noun in capitals is employed followed by an underscore with the letter *n* for noun. For example, the search that belongs to the idiom *to sit in the driver's seat* is therefore *he DRIVE_v* and *she DRIVE_v*. In addition, when looking for *to take somebody under your wing*, *his WING_n* and *her WING_n* are the query. It was considered whether to use the exact meaning of the keyword for the query such as only *his / her head_n* instead of *his / her HEAD_n,* because it is quite certain that results including multiple heads will not refer to a single male or female person. However, singular forms such as *his / her ball_n* can also occur as homonyms; thus, this would mean deleting all these specific occurrences according to their context. Furthermore, the underlying idea of this thesis is that the association of the word in either a more male or female context will influence the comprehension of the idiom. Therefore, it is more important to create an identical testing method as opposed to adding my possibly subjective judgement as to which expressions belong to the query and which do not. The exact queries for the idioms can be found in the tables of this chapter containing the results.

The following Table 27 shows an example of the result of such a query. This particular example shows the search results of *his WING_n*. There are two rows, showing that the lemma search resulted in both *his wing* and *his wings*. Then, the columns per decade indicate the number of instances these phrases were found in the corpus. Furthermore, the total number is easily found on the intersection of the row *total* and column *all*.

HELP	ⓘ	★		ALL	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
1	ⓘ	★	HIS WINGS	878	18	47	52	48	39	30	59	53	88	54	43	25	53	32	26	25	44	59	36	47
2	ⓘ	★	HIS WING	297		21	23	21	10	12	12	8	6	12	14	13	18	12	11	15	25	17	21	26
			TOTAL	1175	18	68	75	69	49	42	71	61	94	66	57	38	71	44	37	40	69	76	57	73

Table 27: Example results COHA query

As already mentioned in chapter 2, the COHA was cleaned up last year; therefore, the template was adjusted to fit this update. The original calculations started at 1810; however, COHA decided not to include the 1810-1819 period, because it was poorly balanced and only contained one million words (*COHA* 2022). Hence, the 1810 bracket of the template was deleted and one for 2010 was added. The time period of 2010-2019 with more than 35 million words is a valuable addition to the corpus (*COHA* 2022) and to the calculations of the template. Furthermore, the total number of words for these time periods was adjusted as they had undergone a change as well. Fortunately, the graphs that portray the results of the Excel tables are connected to the calculations and not to the numbers themselves; thus, a redesign was not needed as they follow the changing input automatically.

Essentially, the graphs that are useful for this study are based on the template of the following Table 28. Additionally, an actual example of this input can be found in Table 29. The first row of input of these tables indicates the decade of the search from 1820 to 2010. The second row shows the number of tokens for the specific search. As discussed in a previous paragraph, this number is the total number of occurrences of the lemmas of the search as seen in Table 27. The third row contains the total number of words in the COHA for that decade. A number which can be found on the website of the COHA. Lastly, the fourth row contains a calculation which divides the number of tokens from the second row through the total words of the third row and multiplies this result by one million thereby computing the normalized frequency.

Excel document	**Entry**
Decade	From 1820 - 2010
Number of tokens	The number of tokens found in the COHA for the search his/her [noun] or he/she [verb]
Total words	The total number of words in the COHA for that decade
Normalized frequency	(Number of tokens / Total words) * 1000000

Table 28: Excel entry

Search (1):	his NOUN			
DECADE	**1820**	**1830**	**1840**	**1850**
# tokens	18	68	75	69
Total wds	6981389	13711287	15807047	16536003
M CxN tokens/m	2,58	4,96	4,74	4,17
Search (2):	her NOUN			
DECADE	**1820**	**1830**	**1840**	**1850**
# tokens	15	34	36	50
Total wds	6981389	13711287	15807047	16536003
F CxN tokens/m	2,15	2,48	2,28	3,02

Table 29: Actual entry example for his / her WING_n

After completing the Excel tables with the number of occurrences throughout the decades, the template provides several graphs of which two will be used to display the results. In the first one, gender bias will be indicated by the tokens of one gender occurring more frequently than the tokens of the other gender per million words. The second graph portrays the difference between the normalized frequencies of the male constructions minus those of the female constructions. Here, there is a male bias when the number stays above one, there is equality when it is exactly one, and there is a female bias when it drops below one. The difference of the occurrences was preferred over the ratio, because the latter gives an error when the number of female tokens is equal to zero. The result is obviously a ratio of zero in that case; however, this does not accurately portray the history of the token as evidenced when compared to the difference graph next to it. As can be observed in Figure 6, there are occurrences of *his tab* in 1830, 1840, and 1900 to 1950 judging from the difference graph, whereas there are none for *her tab*. Consequently, the ratio graph shows a continuous value of zero up to 1960. However, this does not show the most accurate depiction of the situation. Therefore, the decision was taken to only show the normalized frequencies and the difference between the male and the female tokens.

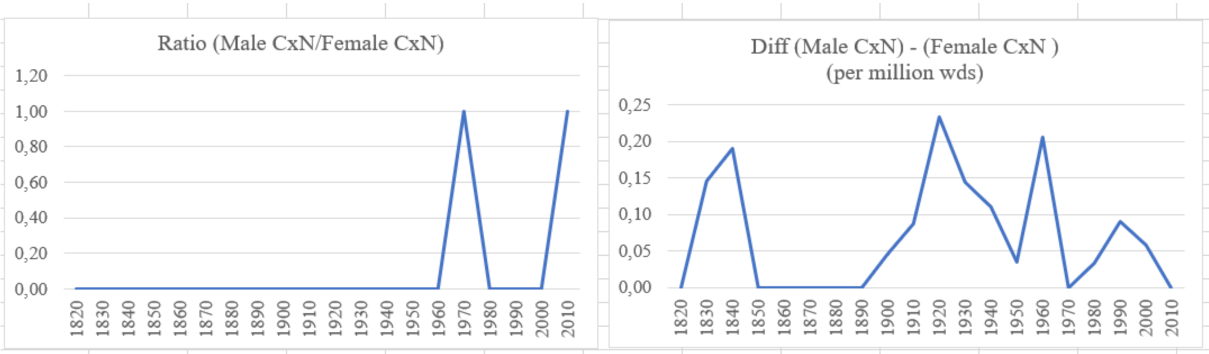

Figure 6: Ratio graph and difference graph his / her TAB_n

4.1. Results

The following sections will each contain a table detailing the idiom that was researched, the keyword that was used in the COHA search, and the total number of occurrences for both the male and the female tokens. Furthermore, the numbers and the graphs that result from them will be presented. Finally, these results are discussed and compared to the hypotheses of this study.

4.1.1. Results dominative paternalism

Idiom	**COHA search**	**He**	**She**
Who wears the trousers/pants	his / her TROUSERS_n / PANTS_N	3019	271
To sit in the driver's seat	he / she DRIVE_v	3581	1190
To call the shots	his / her SHOT_n	408	74
To have the final say	his / her SAY_n	159	33
To run the show	his / her SHOW_n	410	194

Table 30: Results dominative paternalism

The results from the idioms that belong to the category of dominative paternalism show a tendency that was expected. According to the hypothesis of this study, the gender bias present in the keyword of the sentence will influence the meaning of the idiom. The idioms from this category should back the notion that women need to be controlled by men, because they themselves are incompetent (Lomotey and Chachu 2020:72). Hence, it is not surprising that there are more occurrences of his / he + [noun/verb] than there are for her / she + [noun/verb] in the corpus. This is true for all the idioms in this category; however, the relatively low numbers of *to call the shots*, *to have the final say* and *to run the show* emphasise the importance of looking at the diachronic development in order to avoid incorrect conclusions.

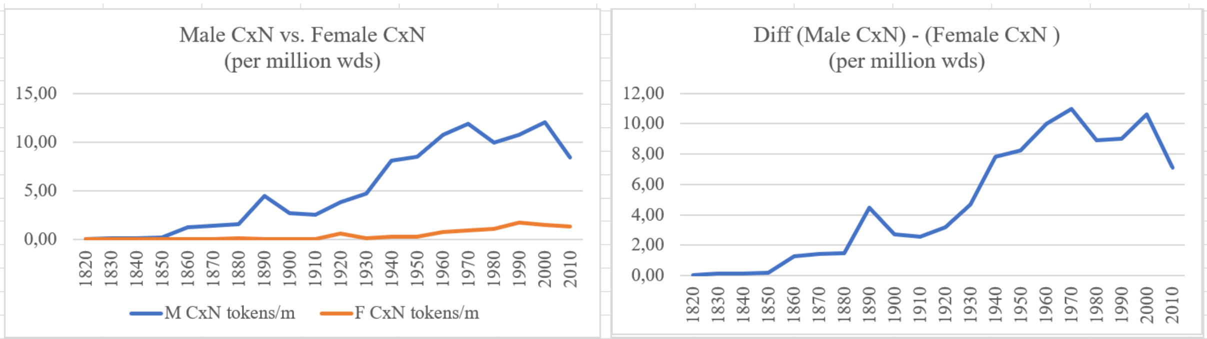

Figure 7: Results his / her TROUSERS_n / PANTS_n

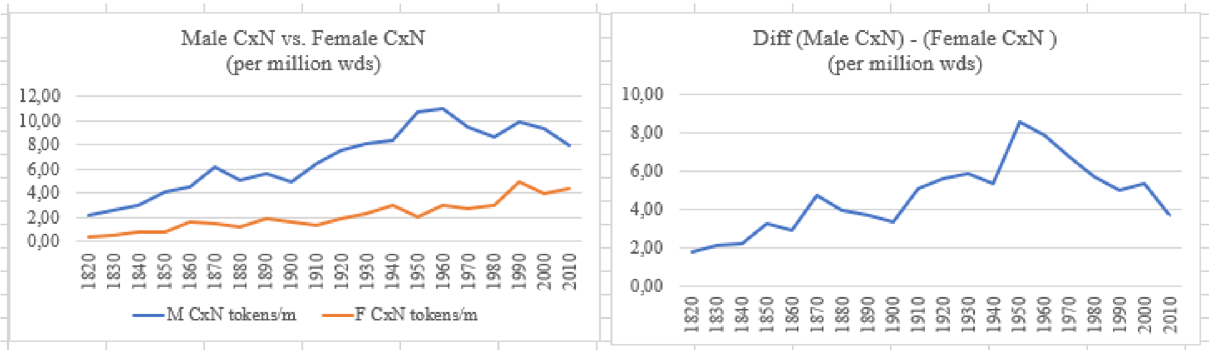

Figure 8: Results he / she DRIVE_v

This tendency for a male bias can best be observed in Figures 6 and 7 above for the noun *trousers / pants* and the verb *drive*. They show that not only the total numbers indicate a male bias, but that also the separate results per decade illustrate this. In the comparison of the normalized frequencies, the blue line never dips below the red line; in other words, the number of male tokens is consistently higher than the number of female tokens Furthermore, the figure displaying the ratio of male tokens divided by female tokens shows that throughout the years, there has been a clear male bias. This bias has a downward trend; however, always remains above one and never shows a female bias. Finally, the difference between normalized frequencies supports these results by showing that the difference between the male normalized tokens minus those of the female normalized tokens remains above zero and therefore demonstrates a male bias.

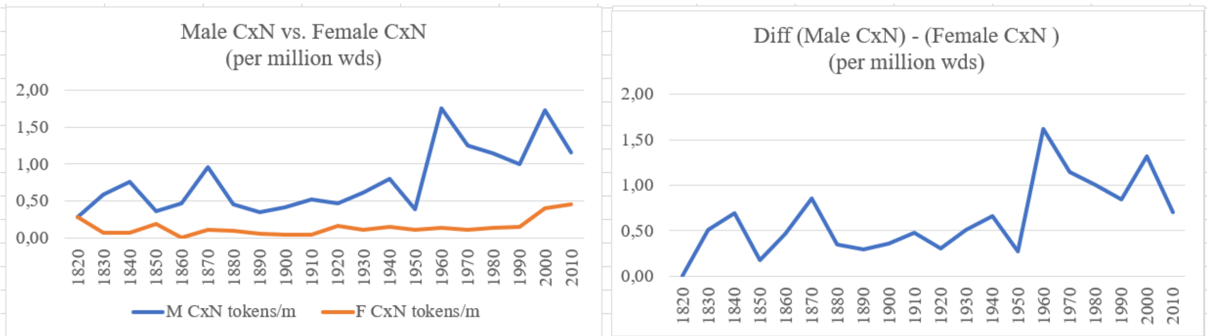

Figure 9: his / her SHOT_n

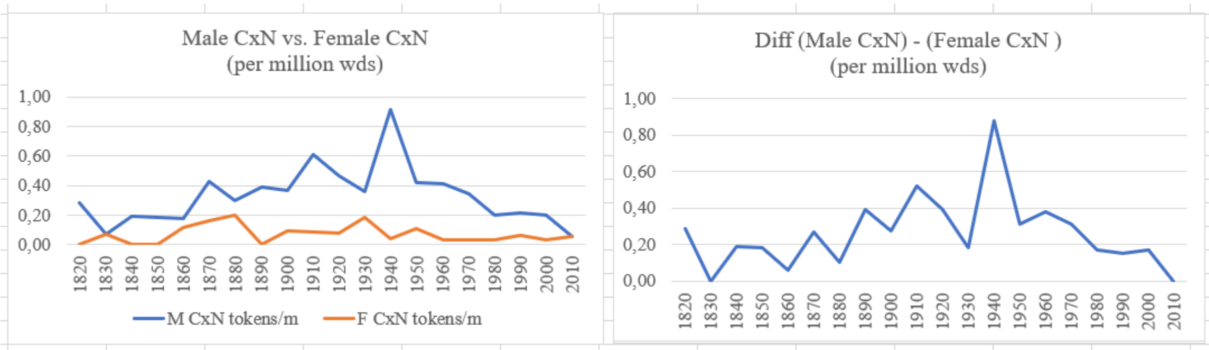

Figure 10: his / her SAY_n

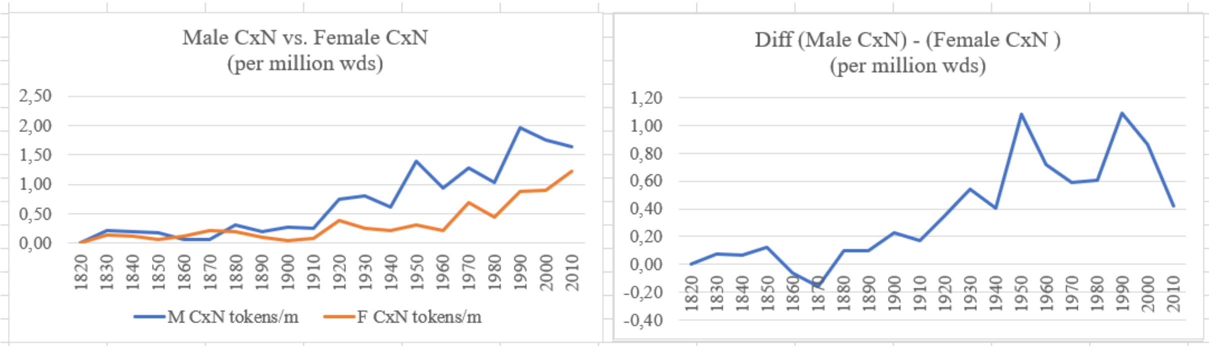

Figure 11: he / she SHOW_v

Even though the other two nouns and single verb from this category clearly obtained more male occurrences than female ones in general, these results are not as decisive as the other two and must therefore be nuanced. Overall, a male bias can be observed in Figures 8-10 above; there are consistently more occurrences of the male version than of the female one, the ratio remains in almost all instances above one, and the difference above zero. There are a few exceptions; however, to understand the importance, or rather insignificance, of this the number of occurrences needs to be taken into account. *His / her SAY_n* shows an equal number in 2010, because there were only two occurrences for both versions. A short, decade-long female bias can be observed in 1870 for *he / she SHOW_v* due to there being one male version and four female versions. In conclusion, a general male bias is present, although it does not have the weight of the previous two.

4.1.2. Results protective paternalism

Idiom	**COHA search**	**He**	**She**
To take someone under your wing	his / her WING_n	1175	601
To pick up the tab	his / her TAB_n	24	19
To stick your neck out	his / her NECK_n	9566	5228
To put your head on the block	his / her HEAD_n	71469	39978
A pillar of strength	his / her STRENGTH_n	3074	1387

Table 31: Results protective paternalism

The idioms representing protective paternalism support the idea that men have more strength and wield more power in society and are therefore responsible for the protection of the women who depend on them (Lomotey and Chachu 2020:72); they are thus expected to show a male bias. When looking at the total numbers, there are more occurrences of combinations with a male pronoun than of those with a female pronoun. At first glance, it can already be said that the findings for *his / her TAB_n* will probably not be usable for a conclusion as the number is quite low.

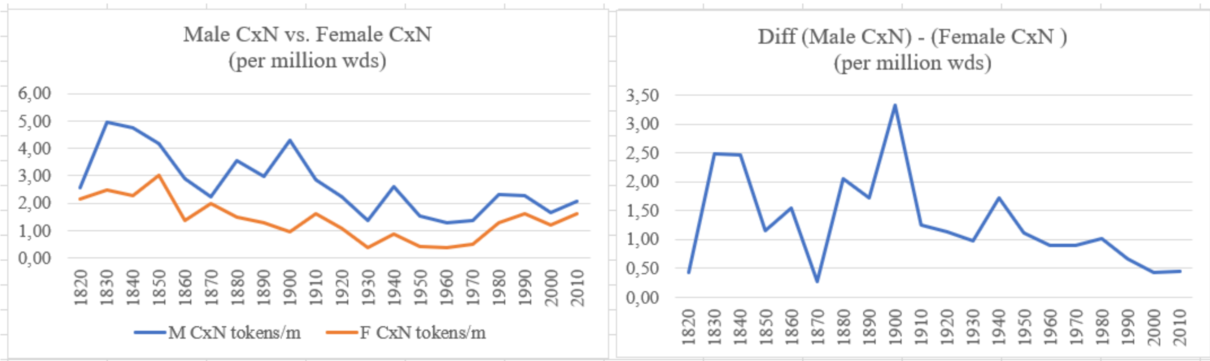

Figure 12: his / her WING_n

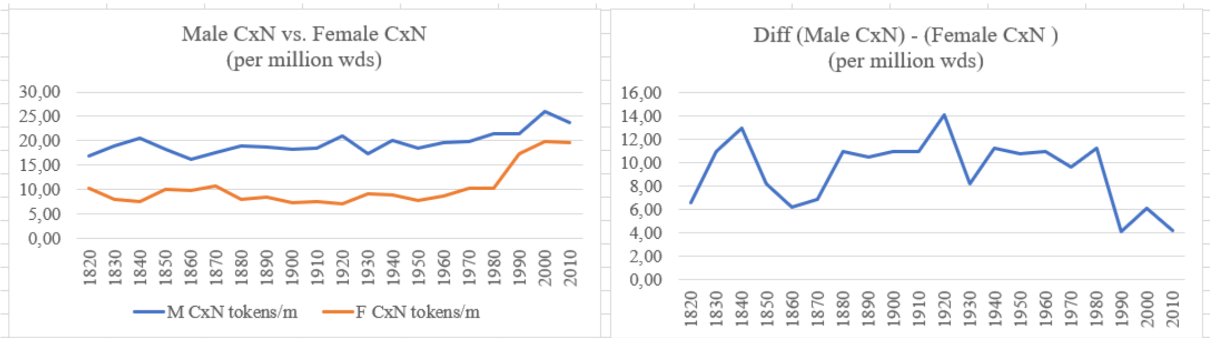

Figure 13: his / her NECK_n

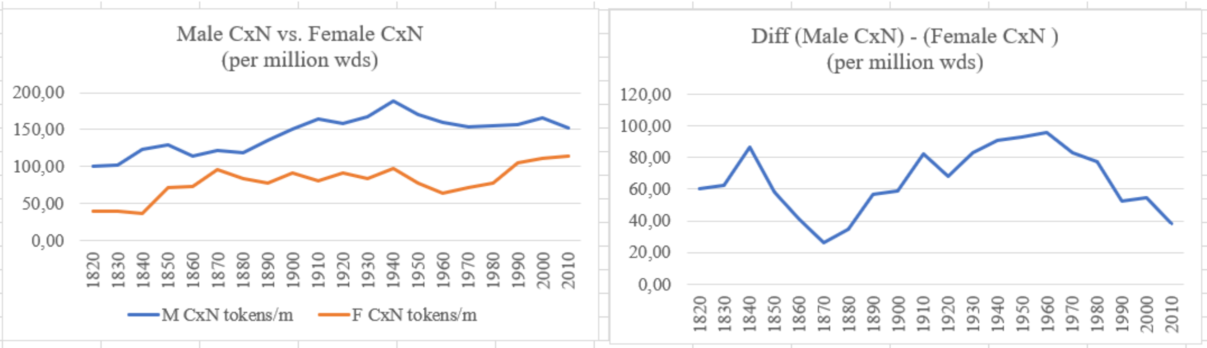

Figure 14: his / her HEAD_n

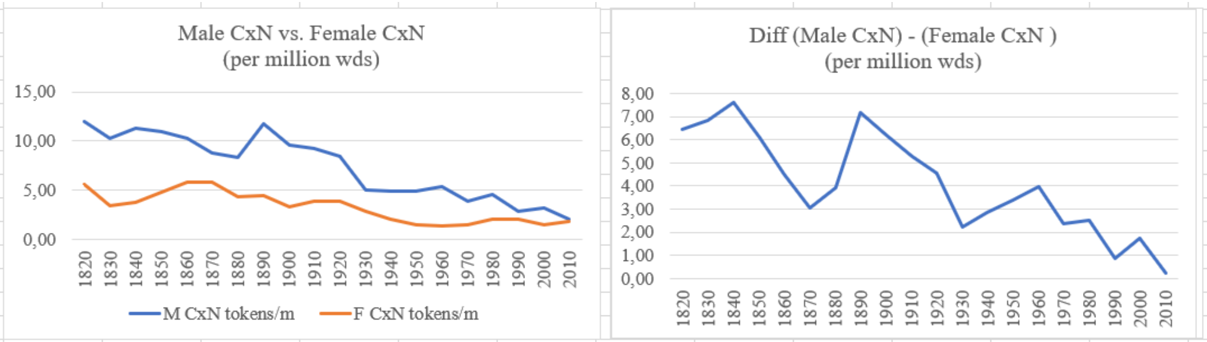

Figure 15: his / her STRENGTH_n

Figure 11-14 endorse the hypothesis that these idioms contain a male bias. Throughout the decades, they consistently show a higher number of male constructions compared to the female constructions. The ratio and the difference also demonstrate this with the values remaining above one and zero respectively. However, it can be observed that these values show a slowly decreasing trend for all four combinations starting around the middle of the last century. *His / her STRENGTH_n* has the strongest decrease ending with only a slight male bias in the 2010's. Nevertheless, it can be deduced that these four idioms diachronically show a male bias.

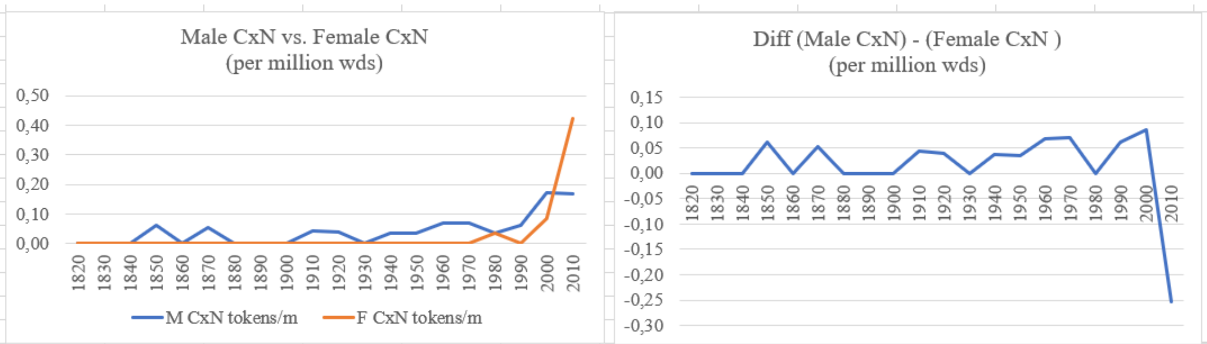

Figure 16: his / her TAB_n

As already predicted, and as can be seen in Figure 15, his / her TAB_n cannot be used to draw any conclusions due to the lack of data. Historically, nothing of significance can be said; in

1980, the first occurrence of *her TAB_n* appears, whereas *his TAB_n* has single appearances throughout the early decades. The last data point of 2010 indicates a female bias with 15 occurrences for the female pronoun and six for the male pronoun. However, these numbers cannot be taken too seriously in a corpus of 35,452,806 words. Furthermore, this female bias is not seen diachronically and is thus not relevant for this study.

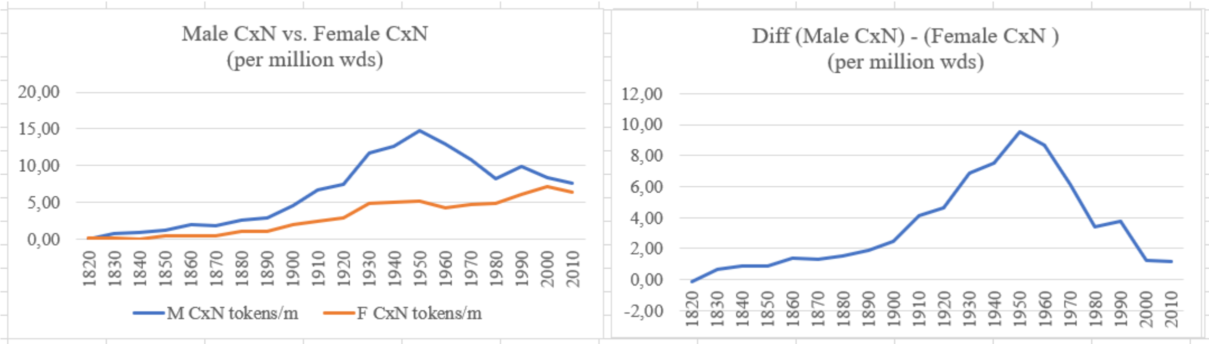

Figure 17: he / she PICK_v up

Alternatively, the keyword *he / she PICK_v up* occurs 3601 times with the male pronoun and 1768 times with the female pronoun. Both these numbers and the resulting graphs in Figure 16 reveal a male bias comparable to the other keywords in this category. The ratio and the difference of this keyword also show a downward tendency ending with a slight male bias in 2010. Therefore, I would argue that this category meets the expected outcome.

4.1.3. Results competitive gender differentiation

Idiom	**COHA search**	**He**	**She**
To be out of someone's league	his / her LEAGUE_n	53	14
To be on the ball	his / her BALL_n	460	85
Smart as a whip	his / her WHIP_n	635	87
Quick on the trigger	his / her TRIGGER_n	37	3
To be up to snuff	his / her SNUFF_n	27	21

Table 32: Results competitive gender differentiation

In this category, the idioms present their subjects as people that have the necessary skills and traits to excel in positions of high power which at the same time also justifies their power (Lomotey and Chachu 2020:72). It follows that the subject of these idioms is expected to be male. The exception to this is of course the first idiom *to be out of someone's league*, because *she* as the inferior person will be out of *his* league and not vice versa. The total numbers support the hypothesis of a male bias; however, this is also the case for the one idiom where a female bias was expected. Similarly to *his/her TAB_n* from the last category, several idioms probably cannot be used to draw conclusions due to their low numbers.

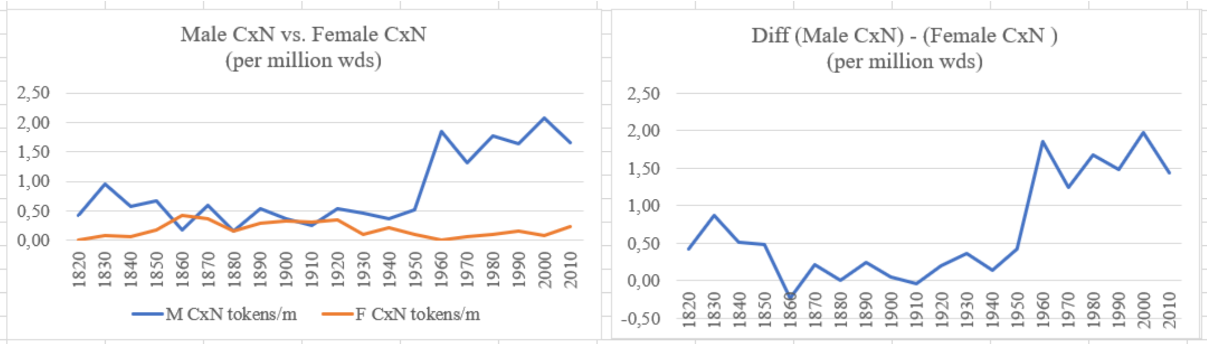

Figure 18: his / her BALL_n

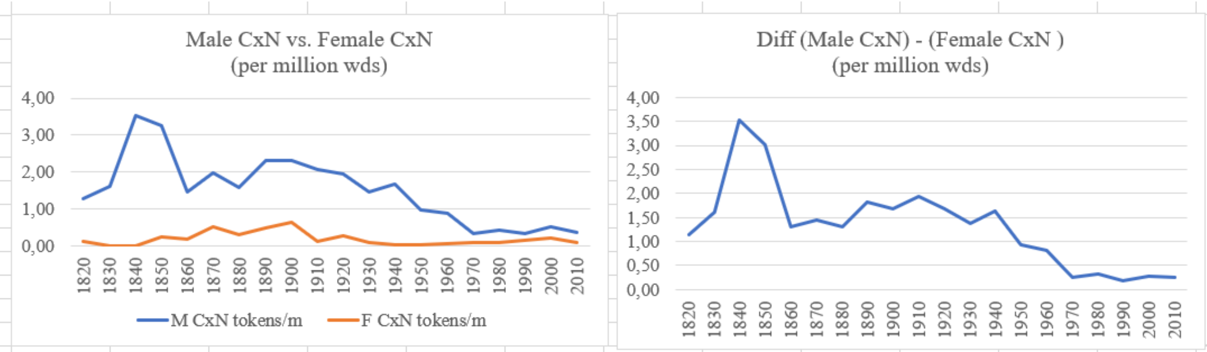

Figure 19: his / her WHIP_n

The two idioms with the highest scores show a general male bias in Figure 17 and 18, although *his / her BALL_n* is not as decisive as *his/her WHIP_n*. The first shows a slight female bias in 1860 and 1910 with respectively three male pronouns versus seven female ones, and six versus seven, while a noticeable male bias steadily increased from the 1940's onwards. I would argue that this increase in numbers and this rise in male bias from the past 80 years weighs more into this study than the first decades with hardly any data. The second displays a clear male bias throughout the years. Mirroring the idioms from the previous category, the bias indicates a downward tendency; however, it remains above one for the ratio and zero for the difference.

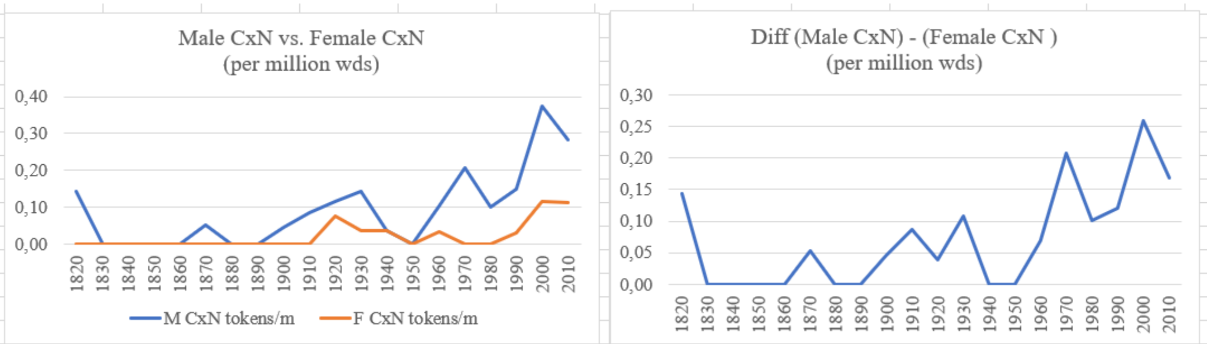

Figure 20: his / her LEAGUE_n

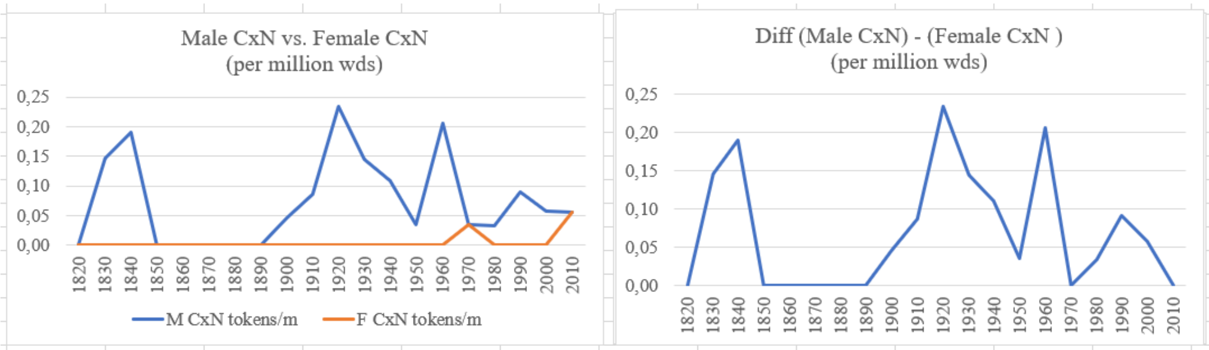

Figure 21: his / her TRIGGER_n

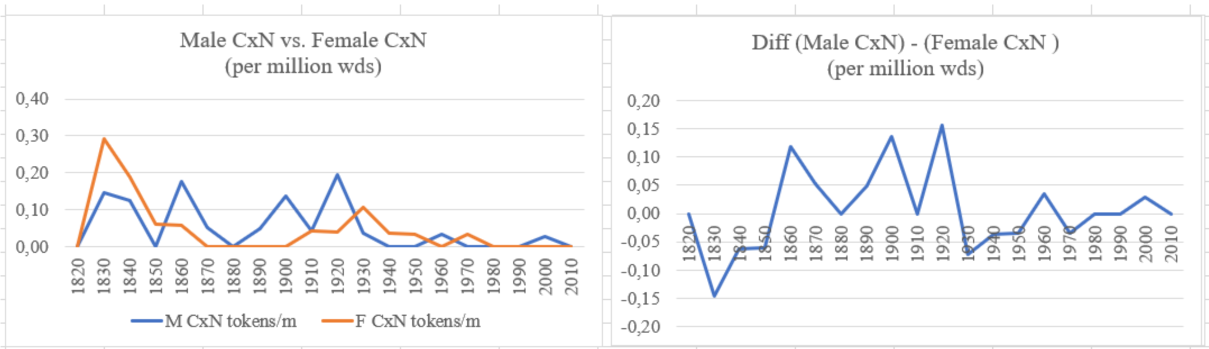

Figure 22: his / her SNUFF_n

The queries of the three idioms from the Figures 20-22 above do not produce enough data to analyse. The highest number of instances for *his / her LEAGUE_n* is in 2000 with 13 male tokens and four female tokens, for *his/her TRIGGER_n* in 2010 with two tokens in both categories, and finally for *his/her SNUFF_n* in 1920 with five male tokens and only one female token. The remaining years show even fewer results.

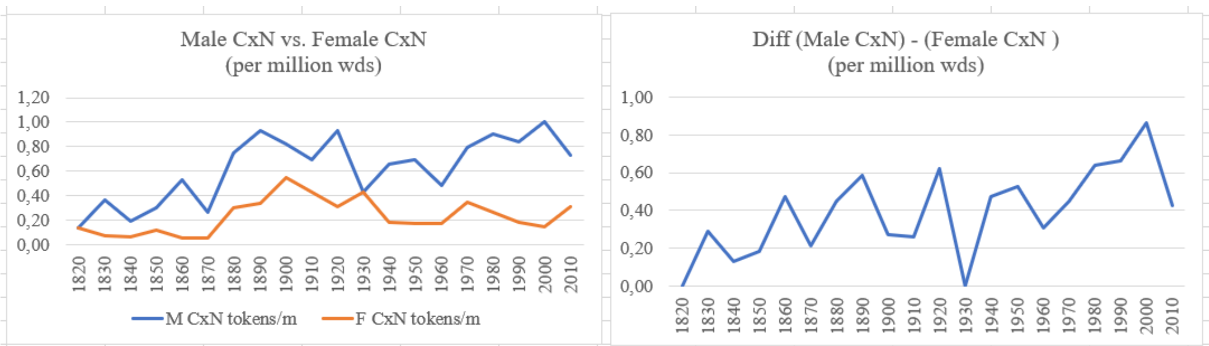

Figure 23: he / she BE_v quick

Unfortunately, there are no alternative searches for *to be out of someone's league* and *to be up to snuff*; however, as an alternative to *his/her TRIGGER_n, he/she BE_v quick* was used as keyword in the COHA search. This query delivers more results with 323 male tokens and 117 female tokens. Here, the graphs in Figure 23 turn out as expected: the occurrences of *he BE_v quick* never dipping below those of *she BE_v quick*, the ratio never dropping below one and the difference never below zero. It must be mentioned that there is one decade with gender equality:

in 1930 there are 12 occurrences for both the male and the female tokens. This, however, is the maximum number of occurrences of *she BE_v quick* seen in 1930 and 1990. The remaining decades show much lower numbers. Therefore, I would claim that this idiom does contain a male bias.

Overall, three out of five idioms that were expected to show a male bias lived up to this expectation. The keywords of the remaining two did not provide enough results to draw a conclusion.

4.1.4. Results complementary gender differentiation

Idiom	COHA search	He	She
To break your back	his / her BACK_n	14695	2066
To cook up a storm	he / she COOK_v	188	303
Pull your weight	his / her WEIGHT_n	1651	658
To be at one’s beck and call	his / her CALL_n	1384	412
To hold one’s tongue	his / her TONGUE_n	3903	2140

Table 33: Results complementary gender differentiation

Contrary to the previous idioms, those of complementary gender differentiation are expected to have female subjects. In this category, women will be ascribed positive character traits that are associated with the way that the gender roles are traditionally divided (Lomotey and Chachu 2020:72). The men’s behaviour is complemented by that of the women and the men are thereby compensated for the characteristics that they lack (Lomotey and Chachu 2020:72). Therefore, an all-female bias is expected, except for *to be at one’s beck and call*; the women that take on the traditional gender role will be at *his call* instead of the other way around. However, when looking at the total number of occurrences, this female bias is only visible in *he / she cook_v*.

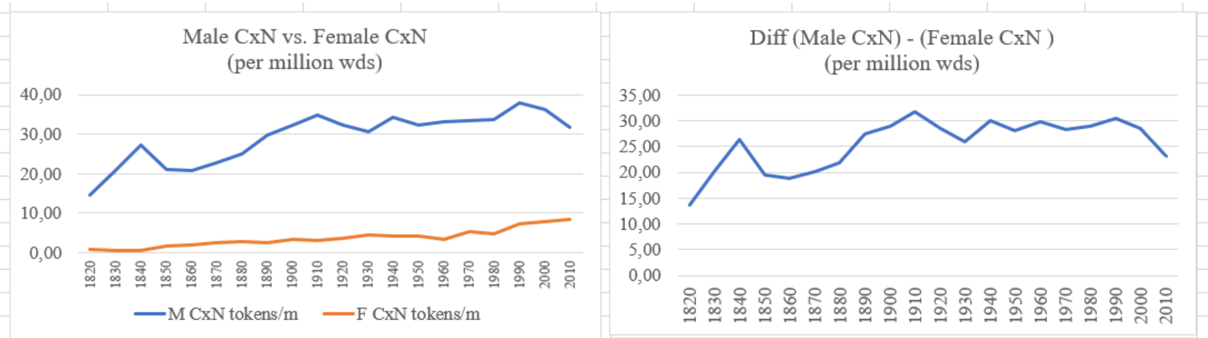

Figure 24: his / her BACK_n

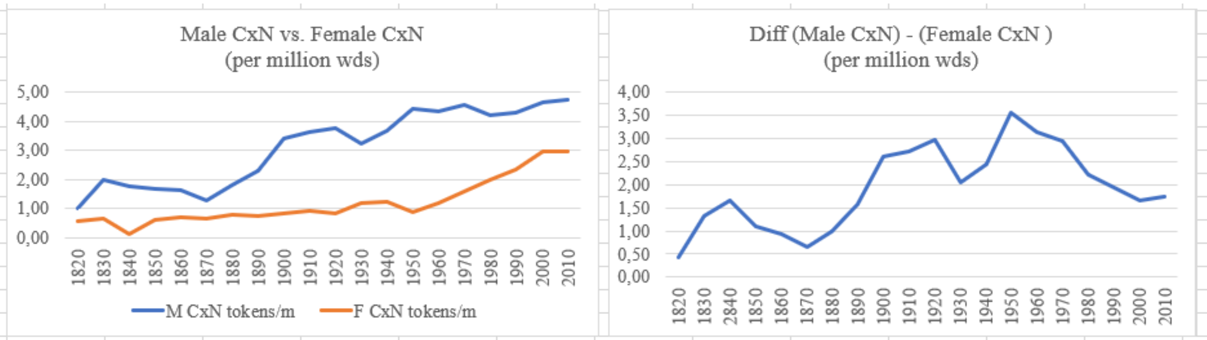

Figure 25: his / her WEIGHT_n

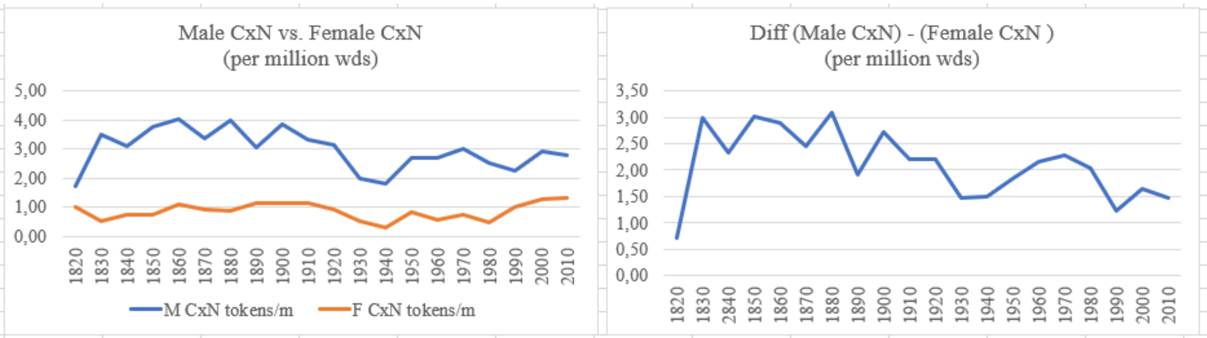

Figure 26: his / her CALL_n

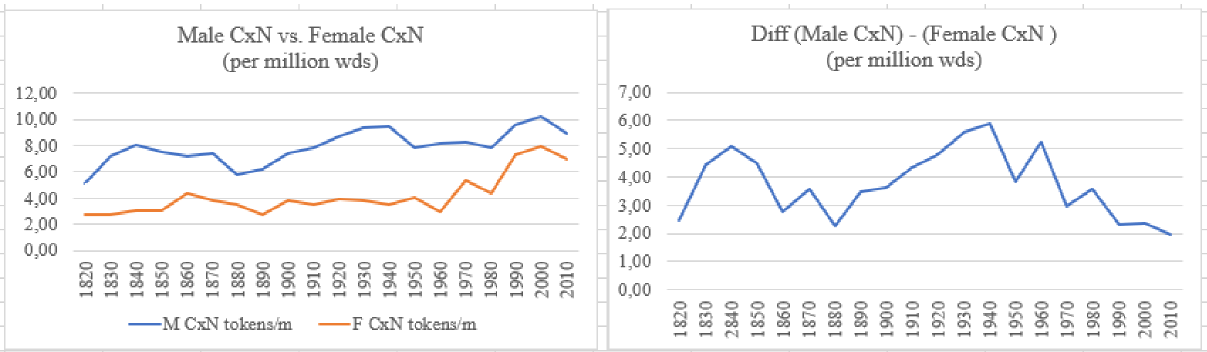

Figure 27: his / her TONGUE_n

These figures 23-26 show a familiar course; they indicate a male bias through the number of male tokens being consistently higher than those of the female tokens, the ratio never reaching one, and the difference never arriving at zero. Furthermore, the ratio and the difference show a downward tendency in the last decades from the middle of the 1900's onwards. This all indicates that the overall male bias reduces throughout the years, but remains present nonetheless.

Figure 28: he / she COOK_v

In figure 27, however, there is a clear female bias. During the first few decades, the numbers are similar, but also relatively low and therefore not conclusive. The 1880 data point shows three occurrences of both male and female tokens; thereafter, the female tokens steadily increase up to a height of 44 tokens, which is double the number of male tokens, in 1990. Therefore, *to cook up a storm* is the only idiom that behaves as predicted by showing a female bias throughout the years.

4.1.5. Results heterosexual hostility

Idiom	**COHA search**	**He**	**She**
Wandering eye	he / she WANDER_v	991	392
Love '-'em and leave '-'em	he / she LEAVE_v	16815	7242
Of easy virtue	his / her VIRTUE_n	445	312
The old ball and chain	his / her CHAIN_n	361	92
To be tied to the apron strings	his / her APRON_n	150	1260

Table 34: Results heterosexual hostility

This category is represented by idioms that show a combination of sex and power: on the one hand, women are only "sexual objects", but on the other hand, they use this sexuality to take away power from men (Lomotey and Chachu 2020:72). The first aspect can be seen in *wandering eye* and *love '-'em and leave '-'em,* while the second feature can be found in *of easy virtue*, *the old ball and chain*, and *to be tied to the apron strings*. The total numbers of the COHA searches confirm the predicted biases: *he WANDER_v*, *he LEAVE_v*, and *his CHAIN_n* obtain more results than the searches for the female equivalent, while *her APRON_n* achievesmore than their male counterparts. *Her VIRTUE_n*, however, appears to have an unexpected slight male bias.

Figure 29: he / she WANDER_v

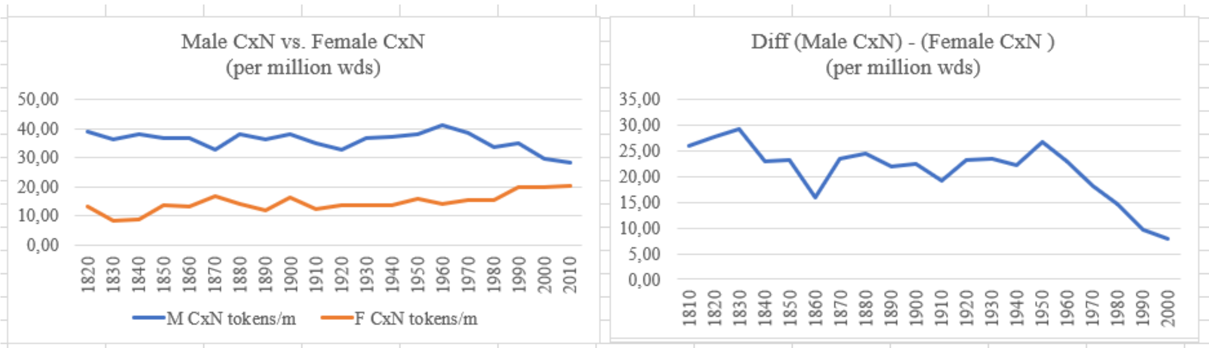

Figure 30: he / she LEAVE_v

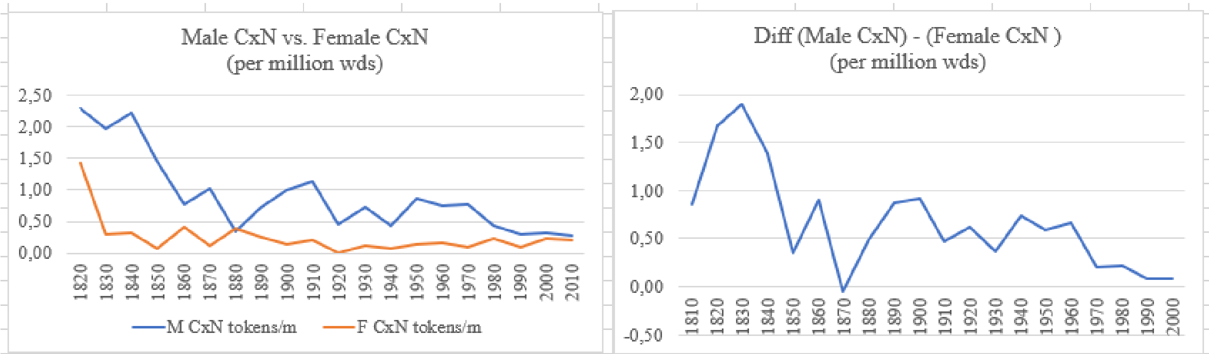

Figure 31: his / her CHAIN_n

The three idioms that were predicted to show a male bias do indeed live up to this expectation as can be seen in the previous figures 28-30. Of the three, *his / her CHAIN_n* has the weakest male bias due to the lower occurrence numbers. In 1880, there is even a female bias due to the number of male tokens, namely seven, being lower than the female tokens, eight. However, it should be noted that this is the data point where two extremes meet; for the male tokens this is the lowest data point, and for the female tokens the highest. The remaining decades show much larger differences as is clearly portrayed in the difference graph. Furthermore, *his / her CHAIN_n* indicates the familiar tendency of the male bias decreasing slowly throughout the decades towards the present time as does *he / she LEAVE_v*. *He / she WANDER_v* goes through a different pattern with large spikes of male bias around 1910, 1940, and 1980.

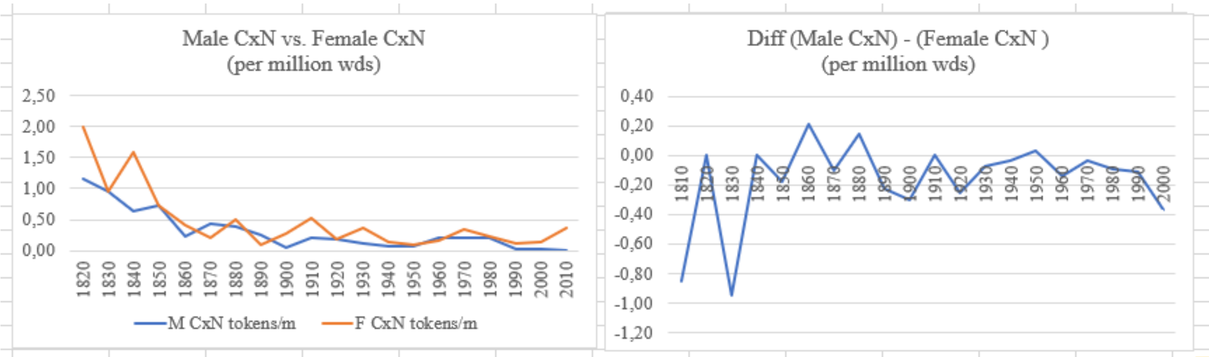

Figure 32: his / her VIRTUE_n

His / her VIRTUE_n in Figure 32 is not as decisive as the previous keywords, and needs a closer look to determine what the graphs indicate. The first decades, the occurrences of both *his* and *her VIRTUE_N* decrease rapidly as does the male bias. After 1930, this bias changes from male to female several times; however, this is to be expected with low occurrence numbers ranging from 4 to 18. Therefore, I would argue that these results are inconclusive.

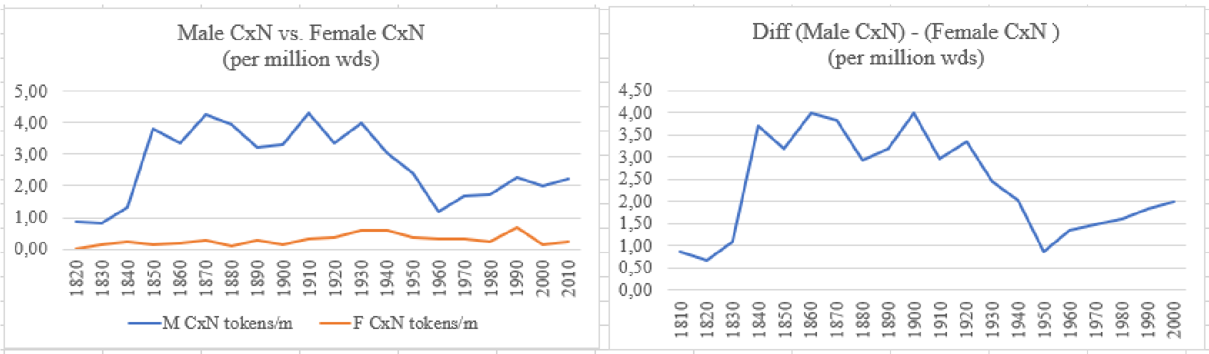

Figure 33: his / her APRON_n

When looking at Figure 33, it is clear that there is a female bias for *his / her APRON_n. Her apron* is consistently more in use than *his apron*; diachronically, there is not one point at which this is not the case. The ratio and the difference graph also confirm this. Thus, is it safe to say that *his / her APRON_n* confirms the hypothesis.

4.1.6. Results heterosexual intimacy

Idiom	**COHA search**	**He**	**She**
To be the apple of my eye	his / her EYE_n	71984	52435
The better / other half	his / her HALF_n	243	72
Trophy wife / husband	his / her TROPHY_n	91	25
To get hitched	he / she HITCH_v	170	45
Match made in heaven	his / her MATCH_n	245	63

Table 35: Results heterosexual intimacy

The final category of idioms reflect the idea that the man will be made “complete” by the woman in the relationship (Lomotey and Chachu 2020:72). Consequently, the idioms that fit into this category focus on the way a woman makes a man’s life better. Thus, a man will see his partner as *the apple of his eye* and *his trophy*. Being with her will make him complete and therefore be *his match made in heaven*. In order to complete him, the woman would need to be *his better half*. Finally, the man will be the one that actively *hitches*, because he is the one in power. Hence, this category again has the expectation of achieving an all-male bias. This expectation is met when looking at the total numbers. *His / her EYE_n* certainly has enough results to draw conclusions with; however, the rest needs a more in-depth analysis.

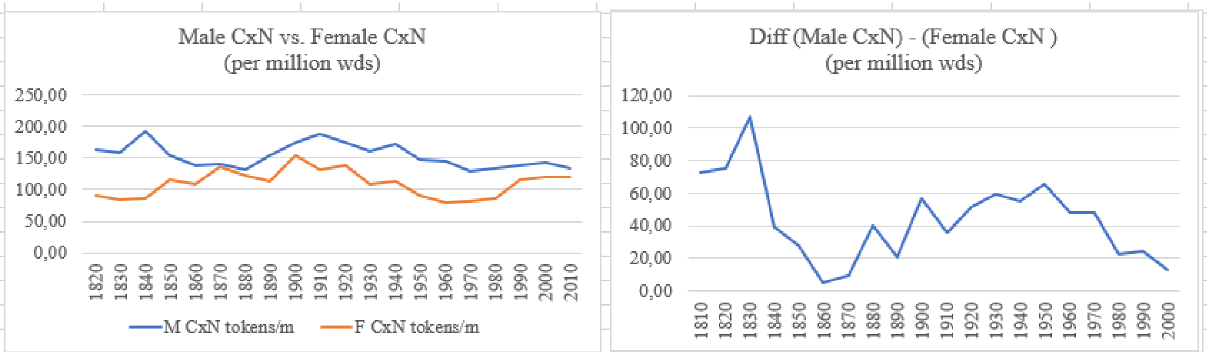

Figure 34: his / her EYE_n

His / her EYE_n in Figure 33 shows a clear male bias throughout the decades. There is a point in time around approximately 1870 – 1880 where the number of female tokens approaches the number of male tokens; however, it never surpasses it. The ratio as well as the difference stays comfortably above the required level of one and zero respectively the entire time.

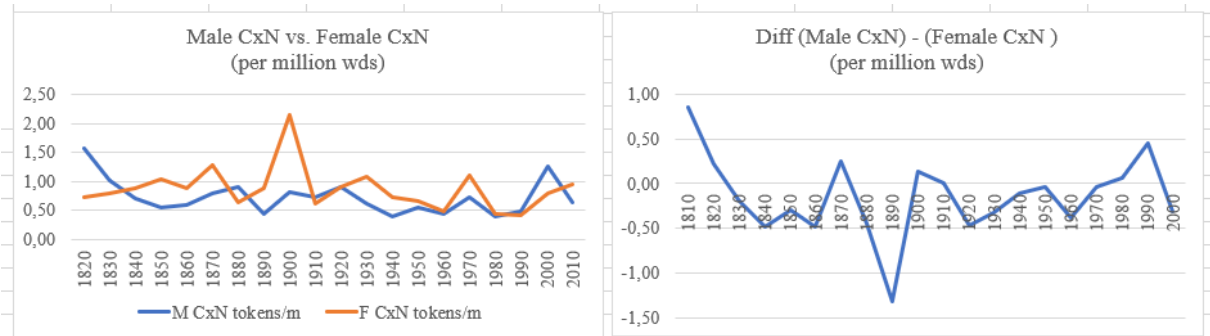

Figure 35: his / her HALF_n

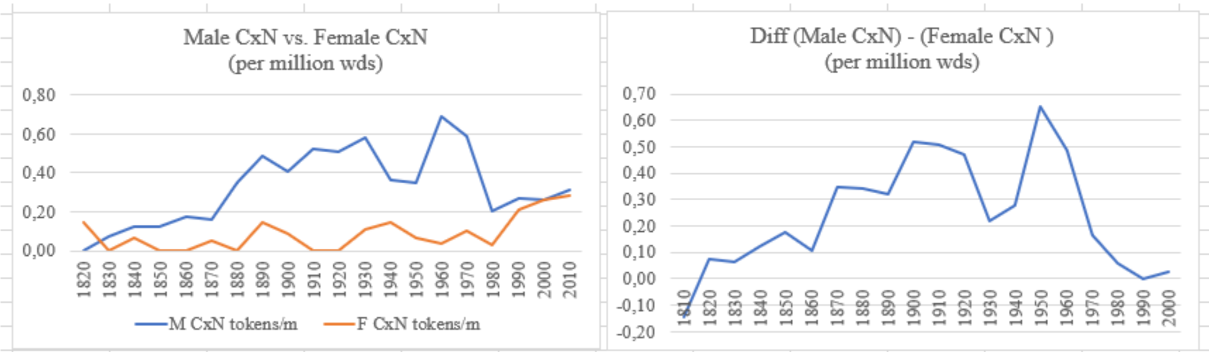

Figure 36: he / she HITCH_v

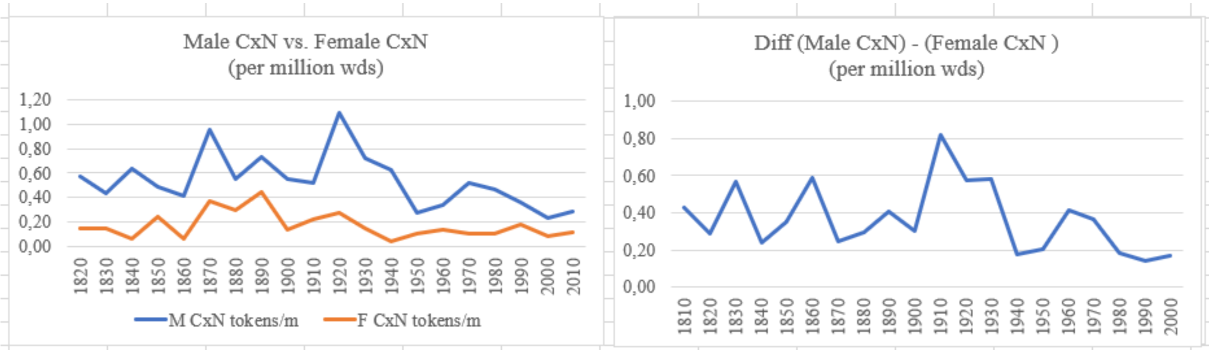

Figure 37: his / her MATCH_n

His / her half_n, *he / she HITCH_v*, and *his / her match_n* in Figures 34-36 all show an overall male bias. First, *his / her half_n* has one moment in 1900 where both the male and the female tokens have 14 occurrences. This causes the ratio to be one and the difference to be zero; thus, a short period of gender equality. However, this is not seen in the remaining decades. Second, *he / she HITCH_v* starts with a female bias, but this is only due to the female category having one token and the male category having zero. The remaining decades show a clear male bias. Please note that the ratio appears to be zero for several decades; however, this is due to the female tokens being zero at that point in time and not due to them being equal to the male tokens. The normalized frequency and the difference graph must therefore always be taken into account additionally. Interestingly, the ratio and the difference decreases sharply after 1990 due to the rapid increase of female tokens. However, this does not bring about a female bias just yet. Finally, *his / her match_n* portrays a clear and uncontested male bias with the blue line never dipping below the red, the ratio never below one, and the difference never below zero. To sum up, these three idioms confirm the hypothesis.

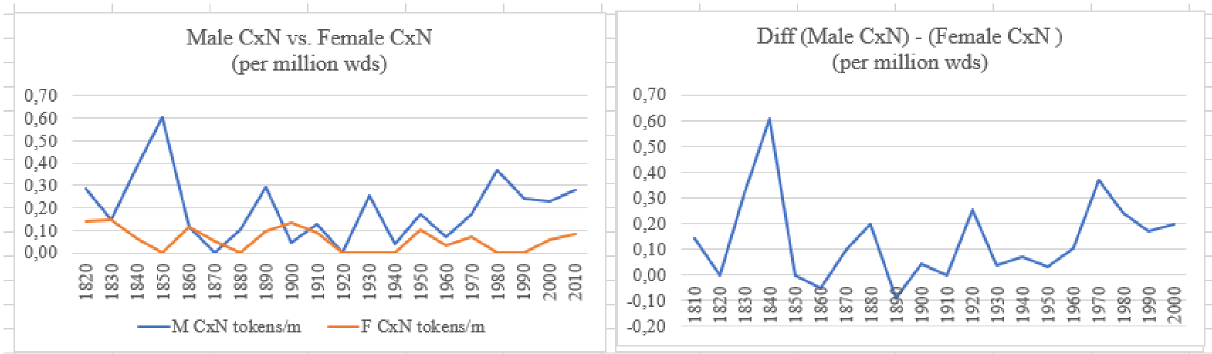

Figure 38: his / her TROPHY_n

Finally, *his / her TROPHY_n* does not have a clear tendency. According to the graphs in Figure 37, there are generally more male tokens than female tokens and a slight, irregular preference for a male bias can be observed. However, when looking at the raw data, the number of occurrences throughout the decades is too low to consider this idiom. Most decades have single digit numbers and oftentimes even no occurrences. Therefore, *his / her TROPHY_n* cannot be used for the conclusion.

4.2. Results overview and conclusion

The following six tables in Table 36 give a clear summary of the data that was discussed in the previous sections. They contain the idiom of the particular type of sexism, the expected gender bias, the actual result, and remarks about the strength of the bias.

Idioms dominative paternalism	**Expectation**	**Result**	**Remarks**
Who wears the trousers/pants	M	M	strong
To sit in the driver's seat	M	M	strong
To call the shots	M	M	weak
To have the final say	M	M	weak
To run the show	M	M	weak

Idioms protective paternalism	**Expectation**	**Result**	**Remarks**
To take someone under your wing	M	M	strong
To pick up the tab	M	M	due to alternative query
To stick your neck out	M	M	strong
To put your head on the block	M	M	strong
A pillar of strength	M	M	strong

Idiom competitive gender differentiation	Expectation	Result	Remarks
To be out of someone’s league	F	X	insufficient data
To be on the ball	M	M	strong
Smart as a whip	M	M	medium
Quick on the trigger	M	M	due to alternative query
To be up to snuff	M	X	insufficient data

Idiom complementary gender differentiation	Expectation	Result	Remarks
To break your back	F	M	strong
To cook up a storm	F	F	strong
Pull your weight	F	M	strong
To be at one’s beck and call	M	M	strong
To hold one’s tongue	F	M	strong

Idiom heterosexual hostility	Expectation	Result	Remarks
Wandering eye	M	M	strong
Love ‘-’em and leave ‘-’em	M	M	strong
Of easy virtue	F	M	weak
The old ball and chain	M	M	weak
To be tied to the apron strings	F	F	strong

Idiom heterosexual intimacy	Expectation	Result	Remarks
To be the apple of my eye	M	M	strong
The better / other half	M	M	weak
Trophy wife / husband	M	X	insufficient data
To get hitched	M	M	strong
Match made in heaven	M	M	strong

Table 36: Overview of the idioms

Overall, the expected gender bias in the idioms follows the expected results. It can be said that on the one hand, the expected male bias in the idioms is supported by data from the Excel template in 21 out of 23 idioms, the other two idioms having insufficient data. On the other hand, female bias is only supported by the data in two out of seven instances, while four idioms show an unexpected male bias, and one has insufficient date. Therefore, I would argue that these 30 idioms representing different types of sexism display a general trend of hidden male bias. However, this conclusion indicates a tendency and does not carry enough weight to support a general claim. In order to do so it would need further research with a focus on investigating a dataset that is more evenly divided, e.g. has an equal division of the expected gender bias of the specific idioms. Furthermore, a larger dataset would provide more insight into this area.

5. Discussion

The main purpose of this thesis was to explore the link between idioms consisting of "outwardly neutral language" (Lomotey and Chachu 2020:70) and their influence on gender bias. Misogynistic ideologies are often hidden in language that appears to be neutral (Lomotey and Chachu 2020:70). Especially in the Western world, where the use of idioms is no longer as prevalent as it used to, it could be assumed that the status quo is evolving (Schipper 2010:19). However, the ideas that were expressed by these idioms have not vanished and are part of an "internalized subconscious legacy" (Schipper 2010:20). Therefore, increasing the awareness of these ideologies in idioms can only be beneficial as they are generally seen as "embodiments of collective wisdom" (Lomotey and Chachu 2020:71).

The results from the questionnaire from this study clearly show the difference of the level of awareness between hostile and benevolent sexism according to the classification of the Ambivalent Sexism Theory by Glick and Fiske (1996, 2010). The results from the Wilcoxon signed-rank test, where groups are tested on related issues, in this case the idioms containing either hostile or benevolent sexism, show that the group as a whole distinguished between the two types of sexism. The difference could not only be found in the group as a whole, but also between the female and the male group, and the native and the non-native group. The benevolent idioms received significantly higher scores than the hostile ones, which confirms the theory that gender bias is more pronounced in the latter. This in turn supports the Ambivalent Sexism Theory by providing evidence of the awareness of the difference between the two types. Whether the participants were actively aware of this difference in their judgements cannot not be determined through this experiment; however, this could be a prospective research opportunity.

Unfortunately, even though it was a primary aim of this study, no pattern could be discovered between the answers of the participants and the corpus research. Furthermore, no such pattern could be detected between the idioms that did have a significant difference. The Mann-Whitney test, comparing the answers of the female and the male group, shows a significant difference in scoring in only six out of 30 idioms or 20%. These six idioms were expected to show a male bias; four of them met this expectation with a strong bias, one showed a weak bias, and the final one did not have a sufficient number of results. Additionally, neither the comparison between the answers of the native and the non-native group could offer more insight. Similarly, a significant difference was only found in 20% of the answers. Again, four idioms met the expected male bias, one showed a weak bias, and the final one resulted in a male bias instead of the expected female one.

It would seem logical to explain inequalities in language through historical developments; however, Criado Perez (2019:8) showed that even new languages show bias. Thus, feminism gaining more traction and the vocabulary around women changing will not counteract the centuries of gender biased language. Therefore, it would appear more obvious to explain bias in language through the theory of linguistic relativity and through neuroscience. Indeed, the first proposes that the language that we speak will influence the way that we think (Lucy 2001:903), while the latter shows that our brains will mirror the type of input that they receive (Rippon 2019:356-357). This is at once a plausible explanation for the continuing sexism in our language and a hopeful outlook towards the future. There are still many people who refuse to acknowledge gender bias in the language; however, according to Pauwels (2003:561), they are indeed aware of it. Pauwels (2003:561) bases this on the growing metalinguistic conduct that supports the increasing level of awareness of gender bias. It follows that the more this awareness grows, the more it will influence our language and subsequently our brains.

Consequently, even though this experiment did not show the expected results, it does not mean that the theorized gender bias in idioms is non-existent. Research by Kim et al. (2019), Lomotey (2019), and Schippers (2010) clearly indicates that the area of proverbs and idioms needs more attention, especially since they are such a relevant part of our daily lives. Therefore, I would advocate for more experiments similar to this thesis to elucidate this area further.

Additionally, even though this was not a research goal, I would claim that the finding that L2 speakers answered the questions without a significant difference from L1 speakers rules out neither Cieślicka's (2006) nor Beck and Weber's (2016) research. Cieślicka (2006:120) reasons that L2 learners will already have encountered the literal meaning of the separate words of an idiom before they happen upon the idiom itself. These already familiar words will then have the strongest activation (Cieślicka 2006:121). In other words, Cieślicka (2006:140) supports the theory of the prioritization of literal meanings during L2 idiom processing. On the other hand, Beck and Weber (2016:12) argue that L2 learners will process the figurative meanings of the language with an ever increasing aptitude as their proficiency in the target language grows. The manner in which idioms are processed will become progressively similar to native speakers (Beck and Weber 2016:12).

At first glance, it would appear that the results could be used to support Beck and Weber's (2016) theory; however, I would argue that Cieślicka's (2006) cannot be discounted. The non-native group of participants definitely processed the idioms in such a way that the resulting numbers did not show any significant difference with those of the native group. Furthermore, their high level of English certainly points towards them processing the idioms in a similar manner. The participants of this study have an average level of almost C1; 91.3% of the participants judge their level to be B1 level or higher, while 66.1% of them even say they are C1 or higher. Nevertheless, the keyword that was used to gather data during the corpus research was based on the occurrences of this particular word in the COHA, whether this was in a figurative or a literal context was not included in the research. The focus was on the occurrence of this specific word in either a male of female biased context. Therefore, the non-native participants could also have focussed, as Cieślicka (2006) theorized, on the literal meaning of the word and answered from their experience with it.

In addition, it could also be argued that the lack of a significant difference between the answers of the native and the non-native group is based in Schipper's (2010:18) discovery that themes and variations of similar proverbs can be found in different cultures and languages. They (2010:299) discovered that proverbs often carry the same ideas and concepts cross-culturally. Thus, the participants could have, either consciously or unconsciously, recognized familiar messages and answered accordingly. Indeed, Schipper's (2010) and Lomotey's (2019) research further supports this by showing that these themes do not have to be presented in an overt manner to be identified. They remain present in contemporary forms that are not as outwardly sexist as they used to (Lomotey 2019:161); however, their ideas remain part of the "internalized subconscious legacy" (Schipper 2010:20). Therefore, it could be claimed that the participants recognized the underlying sexism of the idioms and answered accordingly.

Finally, looking at the results from the COHA study, it can be observed in several idioms that the male bias that was prevalent during the first decades reduces slowly, starting approximately from the middle of the last century. This confirms the findings of Jennejohn, Nelson and Nuñez (2021:796) who argue that gender bias loses its prominence over time in the COHA. In other words, the earlier the decade that is under investigation, the more striking the gender bias (Jennejohn, Nelson and Nuñez 2021:796). These results show that indeed change is occurring overtime. Therefore, it would be intriguing to continue this line of research to discover whether this trend continues and whether this influences the current gender bias both in our languages and in our brains.

6. Conclusion

The research that was described in this thesis was designed to test gender bias in idioms. The goal was to investigate the link between language that is neutral on the outside and contains a hidden gender bias on the inside. This study was based on two methods of quantitative research: historical corpus research and an online questionnaire. The first portrayed the historical development of the idiom's key noun or verb in an Excel template calculating the normalized frequency and the difference between frequencies, while the latter indicated the participants' awareness of gender bias in idioms. These were classified according to the Ambivalent Sexism Theory by Glick and Fiske (1996, 2010). This theory proposes that sexism revolves around three areas, namely paternalism, gender differentiation, and heterosexuality, and that each area contains both a hostile and a benevolent element.

All in all, the results of 229 participants were used for this study: 160 people identified as female, 69 as male. In the female group, there were 66 native speakers, and 94 overall proficient non-native speakers of English; in the male group, this was 24 native speakers to 45 non-native speakers. Due to the data consisting of ordinal, "rank-ordered scales" variables, non-parametric tests were used: the Mann-Whitney compared the data of two independent groups, the Wilcoxon signed-rank test compared the data of two dependent groups, and finally, the Kruskal-Wallis test, investigated the differences between two or more groups.

Firstly, the Mann-Whitney test indicated that the difference between the answers of the group identifying as female and the one identifying as male was only significant in 20% of the idioms; furthermore, the same result appeared for the test comparing the answers of the native and the non-native group. In other words, 80% of the answers of the female and the male group as well as the L1 and the L2 group were similar. Therefore, the research questions that ask whether the results from the participants that identify as female are significantly different from those that identify as male and whether native speakers are more aware of this hidden bias compared to non-native speakers could already be answered with a preliminary negative.

Secondly, the Wilcoxon signed-rank test, showed clearly that the group as a whole distinguished between hostile sexism and benevolent sexism. Not only the group as a whole, also the female and the male group as well as the native and the non-native group showed similar scoring behaviour. The participants showed an awareness of the difference between these two types of sexism by consistently scoring the hostile idioms lower than the benevolent idioms, thereby indicating a more pronounced gender bias in the hostile idioms. Therefore, it can be claimed that there is an overall male bias when comparing the idioms representing hostile sexism with those representing benevolent sexism. It follows these results support the Ambivalent Sexism Theory by Glick and Fiske (1996, 2010) by portraying the awareness of the difference between the two types of sexism. Furthermore, I would claim that native and non-native speakers have the same awareness of sexism. Whether this is conscious or subconscious cannot be determined at this point in time and would possibly need to be researched in the future. Furthermore, no significant difference between the behaviour of the female and the male group was found. These results thus contest the hypothesis that female participants would have significantly different results compared to the male participants.

Finally, the Kruskal-Wallis test was used to examine the differences between the four groups. The results show that there are no significant differences between the answers for the majority, or 88.3%, of the idioms. Therefore, I would argue that the Kruskal-Wallis test combined with the Wilcoxon signed-ranks test does not support the hypothesis that the female group behaves differently from the male group. Indeed, all groups behaved similarly throughout all the tests.

The last research question whether outwardly neutral-looking idioms contain a hidden gender bias was answered through corpus research. The historical development of the idiom's key noun or verb was displayed in a normalized frequency and a difference graph. Generally, a male bias could be observed in the results of 21 out of the 23 idioms that were expected to show a male bias. Female bias was only supported in two out of seven idioms; of the other five, four contained an unexpected male bias and one had insufficient data. Overall, there seemed to be a general trend of male bias in these idioms.

This dataset was not sufficient to support a general claim in this area; however, it does show that there is potential in this line of questioning. Future research should contain a larger dataset and a more equal division of the idioms according to the expected gender bias in order to carry more weight towards a general hypothesis. Furthermore, the connection between the results of Jennejohn, Nelson and Nuñez (2021:796) who find that the COHA's gender bias decreases over time and the observation of this study that male bias in the idioms reduces slowly from the middle of the last century onwards is something that could be worth investigating further. It would be interesting to see whether this trend continues and whether this changes anything in the gender bias of idioms.

List of references

Alatrash, Reem; Schlechtweg, Dominik; Schulte im Walde, Sabine. 2021. CCOHA: Clean corpus of historical American English. https://aclanthology.org/2020.lrec-1.859/ (07.03.2022).

Allport, Gordon W. 1954. The nature of prejudice. Cambridge, Massachusetts: Addison Wesley Publishing Company, Inc.

Bauer, Laurie. 1983. English Word-Formation. https://doi org.uaccess.univie.ac.at/10.1017/CBO9781139165846. (01.03.2022).

Beck, Sara D; Weber, Andrea. 2016. Bilingual and monolingual idiom processing is cut from the same cloth: The role of the L1 in literal and figurative meaning activation. *Frontiers in Psychology*. 7:1-14.

Boers, Frank; Eyckmans, June; Kapper, Jenny; Stengers, Hélène; Demecheleer, Murielle. 2006. Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language teaching research*. 10,3: 245-261.

Boroditsky, Lera. 2000. Metaphoric structuring: understanding time through spatial metaphors. *Cognition*. 75:1-28.

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

Boroditsky, Lera. 2018. Language and the construction of time through space. *Trends in Neuroscience*. 41(10): 651-653.

Brown, James Dean. 2001. Using surveys in language programs. Cambridge: Cambridge University Press

Bolukbasi, Tolga; Chang, Kai-Wei; Zou, James; Saligrama, Venkatesh; Kalai, Adam. 2016. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. https://arxiv.org/pdf/1607.06520.pdf. (09.03.2022).

Cacciari, Cristina; Tabossi, Patrizia. 1988. The comprehension of idioms. *Journal of Memory and Language*. 27: 668-683.

Cameron, Deborah. 1985. Feminism and linguistic theory. London: Macmillan.

Chetty, Priya. 2015. Nominal, ordinal and scale in SPSS. https://www.projectguru.in/nominal ordinal-and-scale-in-spss/. (12.07.2022).

Chomsky, Noam. 1980. Rules and Representations. https://doi.org/10.1017/S0140525X00001515 P. (30.06.2022)

Cieślicka, Anna. 2006. Literal salience in on-line processing in idiomatic expressions by second language learners. *Second Language Research* 22 (2), 115-144.

Corpus of historical American English (COHA). https://www-english-corpora. org.uaccess.univie.ac.at/coha/. (24.10.2021).

Criado Perez, Carline. 2019. Invisible women. Exposing data bias in a world designed for men. London: Penguin Random House UK

Cutting, Cooper J.; Bock, Kathryn. 1997. That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory & Cognition* 25(1): 57-71.

DATAtab. 2021. "Chronbach's Alpha (simply explained) https://www.youtube.com/watch?v=W9uPvAmtTOk. (14.07.2022).

Dörnyei, Zoltán. 2007. Research methods in applied linguistics; quantitative, qualitative, and mixed methodologies. Oxford: Oxford University Press.

Dörnyei, Zoltán; Taguchi, Tatsuya 2010. Questionnaires in second language research. Construction, Administration, Processing. (2nd Edition). New York: Routledge.

Eddington, David. 2015. Statistics for linguists. Newcastle-upon-Tyne: Cambridge Scolars.

Egbert, Jesse; Larsson, Tove; Biber, Douglas. 2020. Doing linguistics with a corpus: Methodological considerations for the everyday user. Cambridge: Cambridge University Press.

Field, Andy. 2009. 3rd edition 2000 original. Discovering statistics using SPSS (and sex and drugs and rock 'n'roll). London: SAGE Publications Ltd

Gibbs, Raymond W.; Nayak, Nandini P. 1989. Psycholinguistic studies on the syntactic behavior of idioms. *Cognitive Psychology*. 21: 100-139.

Gibbs, Raymond W.; Nayak, Nandini P.; Cutting, Cooper. 1989. How to kick the bucket an not decompose: Analyzability and idiom processing. *Journal of Memory and Language*. 28(5): 576-593.

Glen, Stephanie. 2022. Chronbach's Alpha: Definition, interpretation, SPSS. https://www.statisticshowto.com/probability-and-statistics/statistics definitions/cronbachs-alpha-spss/. (14.07.20220).

Glick, Peter; Fiske, Susan T. 1996. The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology.* 70(3), 491-512.

Grande, Todd. 2015. „Normality tests in SPSS". https://youtu.be/2GRZ_d4ftoo. (12.07.2022).

Guttentag, Marcia; Secord, Paul. 1983. Too many women? Beverly Hills: Jossey-Bass.

Hamblin, Jennifer L.; Gibbs, Jr. Raymond W. (1999). Why You Can't Kick the Bucket as You Slowly Die: Verbs in Idiom Comprehension. *Journal of Psycholinguistic Research.* 28(1): 25-39.

IBM. https://www.ibm.com/analytics/spss-statistics-software. (14.07.2022).

Jackendoff, Ray. 1995. "The boundaries of the lexicon". Everaert, Martin; Van der Linden, Erik-Jan; Schenk, André; Schreuder, Rob Lawrence (Eds.). *Idioms: Structural and psychological perspectives*. Hillsdale, NJ: Erlbaum 133-165.

Jennejohn, Matthew; Nelson, Samuel; Nuñez, D. Carolina. (2021). Hidden bias in empirical textualism. *The Georgetown law journal.* 109: 767-811.

Kim, Keum Hyun; Rou, Seung Yoan; Ali, Tengku Intan Marlina Tengku Mohd; Kim, Joseph. 2019. Female stereotyping and gender socialization through proverbs and idioms: a comparative study of Malaysia and Korea. *Asian Women*. 35(3): 25-44.

Larson-Hall, Jenifer. 2016. A guide to doing statistics in second language research using SPSS and R (2nd Edition). New York, NY: Routledge.

Larson-Hall, Jenifer. 2010 Doing statistics in second language research using SPSS. New York, NY: Routledge.

Lee, Tiane L.; Fiske, Susan T.; Glick, Peter. 2010. Next gen ambivalent sexism: Converging correlates, causality in context, and converse causality, an introduction to the special issue. *Sex roles*. 62: 395-404.

Levinson, David; Malone, Martin J. 1980. Toward explaining human culture: A critical review of the findings of worldwide cross-cultural research. Michigan: HRAF Press.

Libben, Maya R.; Titone, Debra A. 2008. The multidetermined nature of idiom processing. *Memory & Cognition*. 36(6): 1103-1121.

Likert, Rensis. 1932. A technique for the measurement of attitudes. *Archives of Psychology*. 22: 5-55.

Lomotey, Benedicta Adokarley. 2019. Towards a sociolinguistic analysis of the current relevance of andocentric proverbs in Peninsular Spanish. https://dx.doi.org/10.4314/ljh.v30i1.7 (18.10.2021).

Lomotey, Benedicata Adokarley; Chachu, Sewoenam. 2020. Gender ideologies and power relations in proverbs: A cross-cultural study. *Journal of Pragmatics*. 168, 69-80.

Low, Graham. 1999. What respondents do with questionnaires: Accounting for incongruity and fluidity. *Applied linguistics.* 20(4): 503-533.

Lucy, John A. 2001. Sapir-Whorf Hypothesis. https://dx.doi.org/10.1016/B978-0-08-097086 8.52017-0. (03.04.2022).

Moon, Rosamund. 1998. Fixed expressions and idioms in English: a corpus based approach. Oxford: Clarendon Press.

Nezhelskaya, Galina; Samarina, Varvara; Shibkova, Oksana. 2018. Expression plane of English idioms in gender semantics studies. https://doi.org/10.1051/shsconf/20185001117. (11.10.2021).

Oakes, Michael. 1998. Statistics for corpus linguistics. Edinburg: Edinburg University Press,

Pauwels, Anne. 2003. "Linguistic sexism and feminist linguistic activism". In: Holmes, Janet; Meyerhoff, Miriam (Eds.). *The handbook of language and gender*. Oxford: Blackwell Publishing, 550-571.

Power Thesaurus. 2021. Power Thesaurus. https://powerthesaurus.org/. (01.03.2022).

Popan, Jason. 2017. Importing data from Google Forms to SPSS (through Excel). https://www.youtube.com/watch?v=IX5yJXjLaGA&t=36s. (11.07.2022).

Quadflieg, Susanne; Macrae, Neil C. 2011. Stereotypes and stereotyping: What's the brain got to do with it? *European review of social psychology.* 22(1): 215-273.

Rippon, Gina. 2019. The gendered brain. The new neuroscience that shatters the myth of the female brain. London: Penguin Random House UK.

Schaul, Brandy. 2015. Report: 92% of online consumers use emoji (infographic). https://www.adweek.com/performance-marketing/report-92-of-online-consumers-use emoji-infographic/. (22.11.2022)

Schippers, Mineke. 2004. Trouw nooit een vrouw met grote voeten. Wereldwijsheid over vrouwen. https://www.universiteitleiden.nl/binaries/content/assets/algemeen/arv lezingen/arvlezing2004.pdf. (06.11.2022).

Schippers, Mineke. 2010. Never marry a woman with big feet. Women in proverbs from around the world. Leiden: Leiden University Press.

Schleef, Erik. 2014. Written surveys and questionnaires in sociolinguistics. https://www.researchgate.net/publication/320078915_Written_surveys_and_question aires_in_sociolinguistics. (20.01.2022).

Schmidt, Silvio; Gull, Sidra; Herrmann, Karl-Heinz; Boehme, Marcus; Irintchev, Andrey; Urbach, Anja; Reichenbach, Jurgen R.; Klingner, Carsten M.; Gaser, Christian; Witte, Otto W. (2020). Experience-dependent structural plasticity in the adult brain: How the learning brain grows. https://doi.org/10.1016/j.neuroimage.2020.117502. (09.03.2022).

Siyanova-Chanturia, Anna; Conklin, Kathy; Schmitt, Norbert. 2011. Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second language research*. 27,2: 251-272.

Sprenger, Simone A. 2003. Fixed expressions and the production of idioms. PhD. https://repository.ubn.ru.nl/handle/2066/63832. (28.04.2022).

Sprenger, Simone A.; Levelt, Willem J.M.; Kempen, Gerard. 2006. Lexical access during the production of idiomatic phrases. *Journal of Memory and Language.* 54: 161-184.

Swinney, David A.; Cutler, Anne. 1979. The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behaviour*. 18: 523-534.

Tabossi, Patrizia; Fanari, Rachele; Wolf, Kinou. 2009. Why are idioms recognized fast?. *Memory & Cognition.* 37(4): 529-540.

Tabossi, Patrizia; Zardon, Francesco. 1995. "The activation of idiomatic meaning". Everaert, Martin; Van der Linden, Erik-Jan; Schenk, André; Schreuder, Rob. Lawrence (Eds.). *Idioms: Structural and psychological perspectives*. Hillsdale, NJ: Erlbaum. 273-282.

Thompson, Ross A.; Nelson, Charles. 2001. Developmental science and the media: Early brain development. *American Psychological Association*. 56(1): 5-15.

Tamir, Diana I.; Thornton, Mark A. 2018. Modelling the predictive social mind. *Trends in Cognitive Sciences.* 22(3): 201-212.

The Free Dictionary. 2021. The Free Dictionary Idiom's Dictionary. https://idioms.thefreedictionary.com/. (01.03.2022).

Van Ginkel, Wendy; Dijkstra, Ton. 2020. The tug of war between an idiom's figurative and literal meanings: Evidence from native and bilingual speakers. *Bilingualism: language and cognition*. 23: 131-147.

Weisser, Martin. 2016. Practical corpus linguistics. An introduction to corpus-based language analysis. Blackwell: Wiley.

Whorf, Benjamin Lee. 2012. Language, thought, and reality, second edition: Selected writings of Benjamin Lee Whorf. Carroll, John B.; Levinson, Stephen C.; Lee, Penny. (Eds.). Cambridge, Mass.: MIT Press.

Zufferey, Sandrine. 2020. Introduction to corpus linguistics. Blackwell: Wiley.

Appendix: Questionnaire

Section 1 of 10

Master Thesis Research on Idioms

Dear participant,

Thank you for taking part in this questionnaire! My name is Fiona Tamminga and I am a master's student at the University of Vienna in the English Language and Linguistics programme. I am investigating the use of idioms for my MA thesis.
lars
Idioms are phrases that mean more than their separate words together. An example would be “to hit the hay”. This does not mean literally hitting the hay, but it means going to sleep. You could say: “I am so tired, I will probably hit the hay early tonight.”

In the next slides, you will be asked to read a story about two characters. After this, you will be asked to fill in a score indicating how much you think certain idioms fit the characters in the text. The responses to this questionnaire will be saved anonymously.

Do not worry if you are not familiar with an idiom. This is not a test so there are no "right" or "wrong" answers. Try to follow your instincts and fill in the score you think is best. It will take you around 15 minutes to complete the whole questionnaire.

I know that this does not reveal much about the actual research, but I hope you can understand that providing more information at this stage would bias the results. If you have any questions and/or you would like more information about my research, please send me an e-mail: fiona.tamminga@gmail.com; I will be happy to tell you more about the study.

By going to the next slide, you will start the questionnaire.

Please create a username under which your results can be saved anonymously. You could take the first two letters of your first name and the first two letters of your last name (mine would be FITA). Or you could use your own invention.

Section 2 of 10

Story Time!

Read the following (love)story of York and Rome.

You will be asked to score idioms referring to Rome and York after this text, so make sure to read carefully.

Section 3 of 10

Meet York and Rome: Teenagers

York and Rome live in the same village and they often see each other at the local café where they go to meet their friends after school. York's group of friends consists mostly of football players and they hold lively discussions about the developments of the football season, dating, and the money they need to buy a car. Rome's group of friends usually talk about their theatre lessons and gossip about classmates and their love lives. Even though their interests are very different, York and Rome have exchanged numbers. York was the one who walked over and asked for Rome's number.

They start sending each other messages. The thing they notice quickly is that York is good at mathematics, while Rome is good at languages. This makes for a good exchange and their friendship grows while they help each other out with school assignments.

After some time, York asks Rome out on a date. York makes sure to pick Rome up by car and to promise to bring Rome back again afterwards. York has organized everything: a romantic restaurant where a special roof top bar offers an amazing view of the sky, a meal with Rome's favourite food, and live music. Afterwards, they go to the cinema and watch a movie that Rome has wanted to see for some time: a rom-com with a happy, sappy ending.

The first date is followed by many more and they soon make their relationship official.

Section 4 of 10

Meet York and Rome: Adults

Rome and York have managed to stay together after finishing high school.

They still live in the same village and they have recently moved in together into a shared apartment. York has managed to get a job as a senior business analyst at a big international financing company. Rome has translated the interest in languages into a job at the local community college, teaching students German and French. York works full-time, while Rome works part-time, freelances as a photographer, and supports York’s career.

One night, after finding out they are having a baby, York takes Rome back to that same roof top bar of their first date. Marriage is proposed and lots of tears flow along with a big “yes” to the question. After the birth of their baby, Rome is the stay-at-home parent who takes care of everything, while York is successfully advancing on the career ladder.

Section 5 of 10

Pause before the questions

Do you remember who York is?

Do you remember who Rome is?

If you would like, you can go back and refresh your memory.

The questions start on the next slide. Please do not go back once you have started with the questions.

Section 6 of 10

Question-time! (1/3)

Please score the idioms; do you feel it fits more with York's character, or with Rome's?

Do not worry if you do not know some of the following idioms. There are no wrong answers so click on the score that you feel is best.

In this relationship:

Who wears the trousers/pants?

*

York

1

2

3

4

5

Rome

Who is more likely to take someone under their wing?

*

York

1

2

3

4

5

Rome

Who sits in the driver's seat?

*

York

1

2

3

4

5

Rome

Who do you think caught the other person's eye first?

*

York

1

2

3

4

5

Rome

Who picks up the tab?

*

York

1

2

3

4

5

Rome

Who is the lovey-dovey person

*

York

1

2

3

4

5

Rome

Who calls the shots?

*

York

1

2

3

4

5

Rome

Who sticks out their neck?

*

York

1

2

3

4

5

Rome

Who goes the extra mile?

*

York

1

2

3

4

5

Rome

Who has the final say?

*

York

1

2

3

4

5

Rome

Who puts their head on the block?

*

York

1

2

3

4

5

Rome

Who is the pillar of strength?

*

York

1

2

3

4

5

Rome

Who prefers to be on the safe side?

York

1

2

3

4

5

Rome

Who runs the show?

*

York

1

2

3

4

5

Rome

Section 7 of 10

Question Time (2/3)

Please score the idioms; do you feel it fits more with York's character, or with Rome's?

Of the two who would you describe as:

Being out of someone's league?

*

York

1

2

3

4

5

Rome

Breaking their back?

*

York

1

2

3

4

5

Rome

Being on the ball?

*

York

1
2
3
4
5
Rome

Cooking up a storm?
*

York
1
2
3
4
5
Rome

Cutting corners?
*

York
1
2
3
4
5
Rome

Being as smart as a whip?
*

York
1
2
3
4
5
Rome

Pulling their weight?

*

York

1

2

3

4

5

Rome

Being at someone's beck and call?

*

York

1

2

3

4

5

Rome

Thinking outside the box?

*

York

1

2

3

4

5

Rome

Being quick on the trigger?

*

York

1

2

3

4

5

Rome

Holding their tongue?

*

York

1

2

3

4

5

Rome

Wearing their heart on their sleeve?

*

York

1

2

3

4

5

Rome

Being up to snuff?

*

York

1

2

3

4

5

Rome

Section 8 of 10

Question Time (3/3 - almost there!)

Please score the idioms; do you feel it fits more with York's character, or with Rome's?

Who do you think:

Has a wandering eye?

*

York

1

2

3

4

5

Rome

Will go the whole nine yards?

*

York

1

2

3

4

5

Rome

Would love 'em and leave 'em while dating?

*

York

1

2

3

4

5

Rome

Is the apple of their partner's eye?

*

York

1

2

3

4

5

Rome

Is the better half in this relationship?

*

York

1

2

3

4

5

Rome

Has fallen head over heels in love?

*

York

1

2

3

4

5

Rome

Is the trophy wife / husband?

*

York

1

2

3

4

5

Rome

Believes in love at first sight?

*

York

1

2

3

4

5

Rome

Is of easy virtue?

*

York

1

2

3

4

5

Rome

Refers to their partner as "the old ball and chain"?

*

York

1

2

3

4

5

Rome

Whispers sweet nothings?

*

York

1

2

3

4

5

Rome

Is tied to the apron strings?

*

York

1

2

3

4

5

Rome

Thinks their partner hung the moon?

*

York

1

2

3

4

5

Rome

Got hitched?

*

York

1

2

3

4

5

Rome

Calls their marriage "a match made in heaven"

*

York

1

2

3

4

5

Rome

Section 9 of 10

Background Information

Thank you so much for answering the previous questions.

All that is left is a quick round of questions to give me some background information about you.

What is your age?

What do you identify as?

Female

Male

Non-binary

Prefer not to

What is your nationality?

Where do you currently live? (country)

What is/are your native language(s)?

What is your level of English?

- Starter (A1)
- Elementary (A2)
- Intermediate (B1)
- Upper Intermediate (B2)
- Expert (C1)
- Mastery (C2)
- Native

Do you speak any other languages? Please list them and use the levels from the previous question to indicate your proficiency (for instance: German - B2)

Section 10 of 10

The end!

Thank you for your participation.

If you have any questions and/or you would like more information about my research, please send me an e-mail: fiona.tamminga@gmail.com.