

ELISABETH KAUKONEN

Revealing the gender in the genderless.  
Estonian gender-marked vocabulary  
and its perceptions





**ELISABETH KAUKONEN**

Revealing the gender in the genderless.  
Estonian gender-marked vocabulary  
and its perceptions



UNIVERSITY OF TARTU

Press

University of Tartu, Institute of Estonian and General Linguistics

Dissertation accepted for the defence of the degree of Doctor of Philosophy on 6 May 2025 by the Committee of the Institute of Estonian and General Linguistics, Faculty of Arts and Humanities, University of Tartu.

Supervisors: Professor Liina Lindström (University of Tartu)

Professor Raili Marling (University of Tartu)

Opponent: Professor Jarmo Jantunen (University of Jyväskylä)

Commencement: 18 June 2025 at 14.15 at Jakobi 2-438, Tartu

This study has been supported by the Graduate School of Linguistics, Philosophy and Semiotics; funded by the European Regional Development Fund (University of Tartu ASTRA Project PER ASPERA).



European Union  
European Regional  
Development Fund



Investing  
in your future

ISSN 1406-1325 (print)

ISBN 978-9916-27-891-8 (print)

ISSN 2806-2450 (pdf)

ISBN 978-9916-27-892-5 (pdf)

Copyright: Elisabeth Kaukonen, 2025

University of Tartu Press

[www.tyk.ee](http://www.tyk.ee)

## ACKNOWLEDGEMENTS

When I started my master's studies in linguistics at the University of Tartu in 2019, having just moved here from Tallinn, I never would have imagined myself completing a PhD in just 6 years after that. As I discovered, if you find a research subject that becomes close to your heart and manage to find terrific supervisors, it is completely doable. Doing a PhD has been a wild ride, accompanied by confusion, self-doubt, terrible anxiety, overthinking, fatigue, making mistakes, endless scrolling of reels on Instagram and watching trash TV to take the edge off the stress, and crying over R. Yet, at the same time, there has been lots and lots of learning and growing, making new friends, bright ideas, nourishing discussions, discoveries, inspiration, fantastic people and places, wine-fueled conversations and laughter, and a feeling that I am doing the right thing. I am so grateful to have embarked on this journey.

First and foremost, I owe my deepest gratitude to my supervisors professor Liina Lindström and professor Raili Marling – my “academic mothers”, who took me under their wing already during my master's studies, and gave me the opportunity to delve into such an unusual topic within the landscape of Estonian linguistics. They offered me the freedom and space to act independently, while also being there to support, listen, encourage, provide feedback, and, when necessary, steer me in the right direction. I deeply admire Raili's sophistication, broad knowledge of feminist theories and issues, vitality as well as her rationality. I am profoundly inspired by Liina's technical and methodological expertise, innovative mindset, directness and authenticity, and her wit. Both have taught me how to brace myself courageously and to stand tall, among countless other important lessons. They will forever be great role models for me, and it has been a privilege to work under their supervision.

To my reviewers, I extend special thanks for providing me with useful feedback, pointing out the uncertainties and deficiencies, and offering constructive, yet helpful criticism. I thank my internal reviewer professor Virve Vihman for her detailed observations and for drawing attention to instances of incomprehensible wording. Furthermore, I am deeply grateful to my oponent, professor Jarmo Jantunen, and professor Pirrko Nuolijärvi for their helpful comments, suggestions, kind and supportive remarks, all of which were invaluable for helping me to improve my work.

I want to thank the people of the Institute of Estonian and General Linguistics – you have all created an incredibly supportive environment for a young scholar to explore and truly understand what doing linguistic research is all about. Thank you, Ann Veismann, professor Petar Kehayov, professor Renate Pajusalu, and other staff of the doctoral program for your encouragement during the seminars and progress reviews. I am also forever grateful to Maarja-Liisa Pilvik, professor Pärtel Lippus, professor Petar Milin from the University of Birmingham and the members of our statistics club, for pulling me out of despair and assisting me with data analysis. I thank Kaidi Lõo for listening to my

experiment ideas and offering useful advice. I thank Polina Oskolskaia for agreeing to collaborate with me on a study and for enriching it with Russian data. I also thank Peeter Tinitis for providing me with data from DIGAR, which formed the basis of the second publication of this dissertation. I would like to thank Arvi Tavast, Margit Langemets and other people of the Institute of the Estonian Language for supporting me, engaging in discussions with me, showing interest in my work and inviting me to speak at your seminars and on your podcast. I would also like to thank dr Frazer Heritage for kindly assisting me with my work on a few occasions. In addition, I extend my gratitude to my mentors at the Radboud University, dr Laura Speed and professor Helen de Hoop. My sholarly visit to Nijmegen was a great experience, through which I gained a lot of new knowledge and was also able to step out of my bubble to see how linguistic research is conducted in an another institution.

During my PhD, I had the great oppurtunity to be involved in research projects, which were funded and supported by influential institutions, such as the CERV-programme of the European Union. I thereby want to thank Praxis Think Tank, who welcomed me into their team for conducting collaborative studies. It was incredibly inspiring and enlightening to work with you.

I would not have managed to go through this journey without the support of my other academic peers, many of whom I am now fortunate to call friends. Lydia Risberg proved to be irreplicable, always ready to provide help and support, listen to my complaints and offer her company for blowing off steam and clearing my head. I will always be inspired by Aet Kuusik, who has expanded my horizons as a feminist scholar and with whom I have had the honor of working together as lecturers, organizers and researchers. Kristel Algvère has helped me to stay sane thanks to her vibrant spirit, and I treasure these spontaneous laughs and chatters we have had in the room 431. I also thank Adele Vaks and Mari-Liis Korkus for their support, as moments of decompression with them were invaluable. They all have played a role in ensuring that I never felt alone.

And finally, an exceptionally deep bow to all my loved ones outside of academia, for believing in me and reminding me of the importance of having a life beyond it. Thanks to Marilyn, Markus, Maria, Eliis, the people in the Messenger chat Namaste Väe Keskus, Hele-Mai, Sanna, Kaarle, Minna, Katarina, the comrades from Ula bar, and many more. I thank my mother Margot for teaching me how to be independent, curious and caring, and my brother Robin for being a great pal and my No. 1 fan. I thank my aunt Terje and her family for the support throughout the years. My greatest gratitude goes to Indrek for being the anchor through it all. I hold the utmost appreciation for his presence, patience, and peacefulness. And I am sorry for being annoying at times (I'm afraid it will happen again though).

Tartu, May 1 2025

# CONTENTS

ACKNOWLEDGEMENTS .....	5
LIST OF PUBLICATIONS .....	8
1. INTRODUCTION.....	9
1.1. Research questions.....	12
1.2. Terminology.....	13
1.3. Structure of the dissertation and the order of the publications.....	15
2. THEORETICAL FRAMEWORK .....	17
2.1. The principles of feminist linguistics.....	17
2.2. Gender in languages.....	20
2.2.1. Generic masculinity .....	22
2.2.2. Gender stereotypes in occupational titles and role nouns .....	23
2.2.3. Feminist language planning .....	26
2.3. Gender in Estonian.....	29
2.3.1. Research on language and gender in Estonian linguistics.....	30
2.3.2. Attitudes towards gender in Estonian language users and communities .....	32
3. METHODS AND DATA.....	35
3.1. Corpus linguistic study .....	35
3.1.1. Used corpora .....	37
3.1.2. Data extraction and analysis.....	38
3.2. Quasi-experimental study: Likert scale survey .....	40
3.2.1. Data collection and analysis.....	41
4. OVERVIEW OF THE PUBLICATIONS.....	43
5. RESULTS .....	47
5.1. Frequencies of gender-marked words .....	47
5.2. Changes in the vocabulary – results of feminist language planning?....	50
5.3. Semantic categories of gender-marked words .....	54
5.4. The impact of language and stereotypes on gender perception .....	60
5.5. Gender bias in a genderless Estonian.....	62
5.6. Limitations and further research .....	70
6. CONCLUSION .....	73
REFERENCES.....	75
SUMMARY IN ESTONIAN .....	91
PUBLICATIONS .....	101
CURRICULUM VITAE .....	196
ELULOOKIRJELDUS.....	199

## LIST OF PUBLICATIONS

**Publication 1 (P1):** Kaukonen, Elisabeth. 2022. Sooliselt markeeritud sõnad eesti spordiuudistes. *Keel ja Kirjandus* 6, 526–545. <https://doi.org/10.54013/kk774a3>

**Publication 2 (P2):** Kaukonen, Elisabeth. 2023. Kes on esinaine? *Keel ja Kirjandus* 3, 328–336. <https://doi.org/10.54013/kk783a4>

**Publication 3 (P3):** Kaukonen, Elisabeth. 2023. Cleaning aunts and police uncles in action. Unveiling gender dynamics in Estonian compound words. *Journal of Estonian and Finno-Ugric Linguistics* 14(3), 137–171. <https://doi.org/10.12697/jeful.2023.14.3.05>

**Publication 4 (P4):** Kaukonen, Elisabeth, Polina Oskolskaia, Liina Lindström & Raili Marling. 2025. Gender, language and labour: Gender perception of Estonian and Russian occupational titles. *Frontiers in Communication* 9. <https://doi.org/10.3389/fcomm.2024.1454022>

# 1. INTRODUCTION

This dissertation examines Estonian gender-marked vocabulary, specifically the frequencies, usage as well as language users' perception of gendered words. Estonian, as a Finno-Ugric language, expresses gender only through lexical rather than grammatical means, and this study specifically focuses on compound words that contain gender markings. The research is grounded on feminist linguistics and employs two different methodological approaches. First, a corpus linguistic method is used to map the existing vocabulary as well as to analyze its contextualized usage. Second, a quasi-experimental method is utilized to explore how language users perceive certain gender-marked words, and whether gendered words used in a neutral way still carry gender bias. In addition, the thesis explores the broader semantic categorization of this vocabulary, specifically words designating occupations.

Feminist linguistics draws on the hypothesis that language reflects and shapes gender-related social circumstances (Cameron 1992; Talbot 1998; Pauwels 2003). This perspective aligns with the principles of critical discourse analysis (see Kress 1989; Fairclough 2001; Blommaert 2005; Halliday 2007), according to which language cannot be considered separately from society. Language both reflects and constructs social practices, understandings, and power relations – including those concerning gender – continuously influencing and reproducing them, while being shaped by them in turn. Within this framework, feminist linguistics critically examines and challenges the assumption of gender as a natural and taken-for-granted phenomenon, questioning linguistic practices also conventionally regarded as such. Research into language and gender aims to “look beyond what appears to be common sense to find not simply what truth might be behind it, but how it came to be common sense” (Eckert & McConnell-Ginet 2003: 9). Drawing on this perspective, this dissertation seeks to uncover how Estonian gender-marked vocabulary conveys broader societal beliefs about women and men. It also examines how gender stereotypes, as oversimplified assumptions derived from these beliefs, and the inequalities they perpetuate become evident through language. Furthermore, an attempt is made to explore how patterns of language use have become “common sense” and what underlying factors contribute to this. Feminist linguistics and critical discourse analysis as theoretical foundations of this study enable the identification of deeper social meanings behind language through a complementary set of methodologies. Quantitative data provided by corpus linguistics and one quasi-experiment helps to find, explain and generalize linguistic patterns, but the theoretical approach traces their broader societal implications, which may often remain hidden when linguistic phenomena are merely collected and described (see McEnery & Wilson 2001: 76–77).

The occupational domain, in particular, has historically been a great source of gender inequalities and stereotyping (Deaux & Lewis 1984), as power dynamics between women and men have been notably uneven in this area. Therefore, building on the idea that language reflects social order, including the power

relations between genders in occupations, the study focuses primarily on occupational titles in order to unpack how language conveys stereotypes and traditional gender roles in Estonian society, as these linguistic units serve as indicators of this bias. Occupational gender stereotypes are a result of long-term gender segregation, which still remains persistent in our society today. For instance, in the Estonian context, according to the latest Estonian census in 2021 (Statistikaamet 2022), over 80% of women work in social welfare, education, and customer service, while men dominate in the primary sector, such as agriculture and forestry, as well as construction. Moreover, Estonian occupational domain is characterized by a high gender pay gap – in 2023, women’s gross hourly wage was 13.1% lower than men’s, with the largest disparities observed in sectors such as wholesale and retail trade, repair of motor vehicles and motorcycles (25.5%), information and communication (25.1%), as well as the finance and insurance sector (Statistikaamet 2024a). This is a clear indicator of gender inequality in the occupational sphere.

Stereotypes stem from a mixture of sources, such as conclusions drawn from real-life interactions, as I will explain in Chapter 2.2.2 of this thesis, alongside prejudices and linguistic influences (Eagly 1987; Eagly & Wood 2012; Gyga et al. 2016). As long as stereotypes persist in a society, they also remain embedded in language, and vice versa. Yet, it is difficult to imagine a language community entirely free of stereotypes – stereotyping is a natural aspect of human behavior, serving as a way of categorizing differences and making better sense of the surrounding world (Lillhannus 2002; Vaidya 2019). Hence, while it is not possible to *eliminate* stereotypes entirely, it is possible to acknowledge them, become more aware of their impact and thereby work to *reduce* their harmful effects. In this process, language plays a crucial role, and this dissertation more broadly aims to shed light on gender stereotypes in Estonian language, and to tackle the biased representations of men and women in language, contributing to greater awareness of inequalities reflected in language, and a better understanding of their presence. The results of this thesis can be considered in efforts to foster a more equal and inclusive language, and by extension, society.

Notably, Estonian has been considered linguistically more gender neutral and less gender-emphasizing than Indo-European languages with grammatical gender, as the expression of gender in Estonian is limited to lexical resources. Consequently, the gender bias in Estonian, like that in other grammatically genderless languages, has been studied to some extent (for Finnish, see Tainio 2006; Pyykkönen et al. 2010; Engelberg 2018; for Finnish and Turkish, see Renström et al. 2022), but not as thoroughly as in grammatical gender languages. With this dissertation, I aim to contribute to the filling of this research gap. The dissertation provides important information about gender perception in a grammatically genderless language such as Estonian, which has received relatively little attention in the landscape of linguistics worldwide. The gender perception of two systematically different languages, Estonian and Russian, is compared in one publication (P4) of the thesis and this crosslinguistic study helps to unveil how

different linguistic structures and societal norms can influence gender perceptions, offering new insights into gender stereotypes in language. It is crucial to study gender bias and stereotypes in a diverse range of languages, as this tests the validity of conclusions about the linguistic representation of gender. The choice of Russian for a contrastive analysis stems from its gender expression system, as the aim was to compare languages with different gender structures. Additionally, Russian is the language with grammatical gender with most extensive contacts with Estonian, although Estonian and Russian speakers live in distinct yet somewhat similar cultural spaces.

The topic of gender representation in Estonian currently is not only largely absent from international linguistic research, but it also remains understudied in Estonian linguistics. There are only some studies, many of them student research – for instance, Kerli Puna’s (2006) master’s thesis on gender-specific vocabulary, Liisi Piits’ (2015) doctoral dissertation about collocations of words referring to people or Lydia Raadik’s (2016) master’s thesis about gender in the example sentences of the Estonian standard dictionary “Õigekeelsussõnaraamat 2013”, all of which I will describe in greater detail in Chapter 2.3.1. Thus, with this dissertation, I hope to contribute to both international and Estonian linguistics, by filling a significant gap and adding to the growing body of literature that seeks to understand interactions between language, society, and gender representation.

My interest in this particular subject gradually developed during my bachelor’s studies. Having been drawn to feminist issues from an early age, I wrote my thesis about gendered representations of women and men in women’s and men’s magazines. When I entered the master’s programme at the University of Tartu in 2019, I had already read Kerli Puna’s work as well as Reet Kasik’s (2006) popular science article about feminist linguistics, and knew that I wanted to explore a similar area. This led me to complete my master’s with a thesis inspired by Puna’s research, concerning gendered vocabulary on the example of the Estonian sports news portals Delfi and ERR. This topic continued to fascinate me, and I felt that there was much more to discover regarding gender in Estonian. At that time, no other Estonian linguists were studying gender in the Estonian language. All this motivated me to pursue a PhD. My aim was to expand the topic of my master’s thesis. Initially, I planned to combine corpus linguistics with experimental methods. I intended to first analyze gender-marked vocabulary in the Estonian National Corpus of 2019, and then design at least two experiments to examine perceptions of gender in various gender-marked words, as well as the social positions and characteristics attributed to individuals referred to by these words. Since the beginning of my doctoral research, newer corpora have been generated (in 2021 and 2023), and it became clear that time constraints would limit my ability to conduct more than one experiment in addition to corpus studies. As a result, the dissertation now includes three publications that rely primarily on corpus linguistics, engaging with corpus-assisted discourse studies (CADS), and one publication that engages with experimental linguistics, leaving ample material for future research. Overall, studying this topic is both necessary and relevant, as it reflects broader issues of social equity. Although Estonian is grammatically

genderless, the society has room for improvements regarding gender equality – that can be illustrated by gender segregation in employment and gender pay gap described above as well as Estonia’s current gender equality index, which is 10 points below the EU average (EIGE 2024).

## 1.1. Research questions

The first aim of this dissertation is simply to map the existing gender-marked vocabulary in Estonian, by examining and understanding which words exist, what they denote, how frequently they are used and in what contexts. Additionally, the dissertation seeks to analyze how certain gender-marked words are perceived. The broader aim is to examine what social information about the genders and their positions is conveyed by the existing vocabulary.

Another goal is also to argue, based on the findings, how Estonian could be positioned within the framework of feminist language planning. There have been no discussions on this linguistic activism in Estonia. However, feminist language planning has been implemented in many other linguistic communities (see Chapter 2.2.3), since it plays a role in promoting gender equality and reflecting evolving perceptions of gender norms. Estonia can benefit from these international parallels. Therefore, I seek to understand whether and how Estonian adapts to contemporary societal changes. In order to draw potential conclusions about this, I will examine changes in the vocabulary – specifically, differences in the usage frequencies of gender-marked nouns in corpora over the past four decades – and its changing usage with the word *esinaine* (‘chairwoman’) as an example. The choice of this time period for analyzing changes is determined, on the one hand, by the availability of language corpora<sup>1</sup>, and on the other, by the significant societal and value system changes that included the increasing prominence of gender-related issues in Estonia since the restoration of independence in 1991. The decision to specifically examine gender-marked nouns was influenced by earlier work of Kerli Puna (2006), who similarly focuses on nouns referring to people that include gender-specific parts. The more precise research questions of the dissertation are as follows:

1. What gender-marked nouns referring to people occur in Estonian and what is the ratio of words referring to women and men?
2. Have the frequencies and usage patterns of gender-marked nouns changed over time? To what extent do shifts in the usage reflect evolving societal gender norms and align with the principles of feminist language planning?
3. Which semantic categories are represented by Estonian gender-marked nouns referring to people?

---

<sup>1</sup> The earliest texts in the Estonian National Corpus date from the 1990s, while earlier texts are available in the database of digitized newspapers (DIGAR).

4. Which occupations are reflected in the gendered vocabulary marking occupations?
5. How do language users perceive occupational titles in Estonian and is this perception influenced more by language or stereotypes? How does this differ from the perception of occupational titles in Russian?
6. Does the usage of gendered words reveal the positions of women and men in the society and how does Estonian gender-marked vocabulary convey gender stereotypes and biased representations of men and women in general?

These research questions are addressed in the four publications of this thesis: questions 1, 3 and 4 are covered in the publications P1 and P3; question 2 is addressed in P1 and P2; question 5 in the P4; and finally, conclusions drawn from all the publications correspond to question 6.

## 1.2. Terminology

For the definition of the terms used in this doctoral thesis, I have relied on various sources and my own interpretations of these terms that have developed during the research process. In this section, I will explain the core concepts used and terminological choices made throughout the dissertation.

**Sex/gender** – in gender research, a distinction has been made between the terms ‘sex’, referring to biological differences, and ‘gender’, referring to socially constructed characteristics attributed to individuals of different sexes (see Talbot 1998; Germon 2009)<sup>2</sup>. In linguistics, the term gender is more commonly used – for instance, Stahlberg et al. (2007) use ‘sex’ to refer to the extralinguistic category and ‘gender’ to the linguistic category<sup>3</sup>. In this dissertation, I use the term ‘gender’, as it encompasses both the linguistic category and the societal beliefs associated with individuals of different sexes that these categories mark. While there exists a variety of gender identities, my work focuses on women and men, since Estonian vocabulary includes distinct markers for these genders, conveying the dynamics and unequal power relations between them.

**Feminism** – feminism is, as explained by Bucholtz (2014: 23), a “diverse and sometimes conflicting set of theoretical, methodological, and political perspectives that have in common a commitment to understanding and challenging social

---

<sup>2</sup> Today, the distinction between sex and gender is considered oversimplified, as both categories are shaped by an interplay of social, political, cultural, and economic factors within a given society (de Lauretis 1987; Butler 1990).

<sup>3</sup> In Estonian, the distinction between *bioloogiline sugu* ‘sex’ and *sotsiaalne sugu* or *soolitus* ‘gender’, may not be as widely recognized (see Viik 2015), and term *sugu* is more commonly used as an umbrella term, encompassing both, depending on the context.

inequalities related to gender and sexuality.” Thus, feminism addresses issues related to the unequal treatment of individuals in positions of lesser power, including those belonging to gender groups or sexual identities that deviate from what is considered a norm in the society (see Motschenbacher 2022). While this dissertation focuses on the inequalities between women and men as reflected in language, a variety of studies in Estonian also address other concerns of feminism, for instance, the treatment and representation of LGBTQIA+ communities and other gender identities beyond the male-female binary in Estonian language (see Toots 2023; Kuusik 2024; Põldsam 2024).

**Sexism** – sexism is a set of attitudes and behaviors that express a negative or stereotypical attitude towards an individual on the basis of their gender. When the term ‘sexism’ emerged in the women’s liberation movements of the 1960s, it specifically referred to discriminatory behavior towards women, highlighting their marginalization in the society. Today, the term is more broadly used to refer to any form of gender-based discrimination, including against men (Manne 2018; Slovská 2022). Frye (1983: 19) explains that sexism occurs in a situation where the gender of a person is made relevant – often in contexts when it is, in fact, irrelevant – and making distinctions based on their gender “reinforces the patterns which make it relevant.”

**Gender stereotypes** – gender stereotypes can be described as beliefs about the characteristics, roles or attributes associated with the representatives of different genders (Vaidya 2019). According to Patricia Niedzwiecki (1993: 15), while sexism is related to stereotypes, stereotypes are related to norms. That is, stereotypes arise from generalizations about which roles and characteristics are considered normal for representatives of different genders, while sexism occurs due to gender-related differentiation that stems from these stereotypes.

**Generic masculinity** – a phenomenon known as generic masculinity refers to language units explicitly marking the male gender that are used to denote a person of any gender, or in contexts where the gender of the referent is unknown. These units are referred to as ‘generic masculines’, ‘male generics’ (in the case of languages without grammatical gender), or ‘androcentrics’. Thus, the meaning of words indicating the male gender has expanded, and they are used in a generic way, in addition to specifically referring to men (Cooper 1989; Pauwels 2003; Stahlberg et al. 2007: 169; Hellinger & Bußmann 2015: 9). While ‘generic masculinity’ is perhaps more often associated with grammatical gender languages, I employ this term, as it is, from my experience, recognized more commonly. Moreover, ‘generic masculinity’ denotes a phenomenon that is also present in vocabulary and, therefore, in grammatically genderless languages, highlights the underlying ideological process that operates similarly across languages. I use the term ‘masculinity’ to refer to words that convey the male gender and therefore also a masculine meaning.

**Discourse** – in this dissertation, discourse is understood through Norman Fairclough (2001) for whom the term, in simple terms, denotes language “as a form of social practice”, and refers to the “whole process of social interaction of which a text is just a part” (Fairclough 2001: 18, 20). Discourse, therefore, integrates language with actions, interactions, beliefs, values, attitudes etc., to convey a specific social message underlying language use (Gee 2005: 21). In this context, discourse reflects the social processes and patterns related to gender that language both implicitly and explicitly communicates, while also participating in their recreation. For the purposes of this study, I do not treat discourse primarily as language in use, but rather as a way in which units of language (i.e., vocabulary) reflect and reinforce social structures and power imbalance regarding gender. That is, I employ discourse and discourse analysis as a theoretical basis of the analysis. Thus, this study does not aim to detect specific discourses in a text. Instead, by analyzing the use of gender-marked words in particular contexts, I seek to uncover the meanings behind their usage, treating language as a lens that reveals underlying social positions of genders.

**Language corpus** – a language corpus is a database consisting of a large collection of digitized texts, which are authentic (i.e., produced with the purpose of communication), representative of a language or language variety, and annotated (i.e., containing information about metalinguistic aspects, linguistic properties of utterances as well as demographic information about the texts) (Stefanowitsch 2020: 22–23). Language corpora are compiled and processed primarily for conducting linguistic research.

**Compound word** – a word that is made up of two or more lexemes (e.g., nouns) that together create a new word with a separate meaning (Bauer 2006). Estonian language incorporates a lot of compounding, which is a frequently used method of word formation (Kasik 2015). The second part (i.e., base form) of a compound word in Estonian denotes the main meaning of a word, whilst the first part (initial form), often in a genitive case, describes its type (*spordimees* = *spordi* ‘sports’ SG.GEN + *mees* ‘man’ SG.NOM, or *meessportlane* = *mees* ‘male’ SG.NOM + *sportlane* ‘athlete’ SG.NOM).

**Gender marking** – I refer to gender marking as an element of a compound word that denotes a gender-specific meaning, either in the initial form or the base form.

### 1.3. Structure of the dissertation and the order of the publications

The dissertation consists of the introductory overview chapter and four research publications, published in 2022–2025. In the following chapters of the overview, I will firstly develop the theoretical framework underlying these studies. The methods and data section is divided into two parts. First, I introduce the language

corpora I used for detecting the vocabulary as well as the steps and methods employed to analyze the findings. Second, I describe the design of the survey conducted with language users, including data collection and the statistical methods used for analysis. Then, I summarize each of the publications. In the results section, I describe the main results of the dissertation, propose answers to the research questions and conclude the dissertation by discussing the limitations of this study and offering suggestions for future research.

The order of the publications is as follows: “Gender marked vocabulary on the example of Estonian sports news” (P1), “Who is chairwoman?” (P2), “Cleaning aunts and police uncles in action. Unveiling gender dynamics in Estonian compound words” (P3) and “Gender, language and labour. Gender perception of Estonian and Russian occupational titles” (P4). The publications are listed in the order in which they were written and published. P4, a cross-linguistic study of Estonian and Russian, is co-authored with Polina Oskolskaia, who conducted and analyzed the Russian survey, and my supervisors Liina Lindström and Raili Marling. My contribution involved planning the Estonian survey, conducting and analyzing the results as well as writing the manuscript. The order of the publications also reflects different stages of this research: they begin with corpus studies mapping vocabulary and usage, and conclude with a comparative analysis of linguistic stereotypes in the perspectives of language users.

## 2. THEORETICAL FRAMEWORK

In the following chapter I will outline the theoretical groundwork and key studies that have informed the writing of this thesis. The theoretical core of this study lies mainly in feminist approaches to language analysis, which view language as a mirror of the social positions and roles assigned to the representatives of different genders. In this context, feminist linguistics, based on Cameron (1992), is also a criticism of the normalized assumptions about genders that are embedded in language. Accordingly, while the study focuses on the *What is there?*, by documenting the existing vocabulary, it also aims to address the *Why is it there?*, by drawing socially critical conclusions from the findings based on the theories of sexism and gender stereotypes in language. The concepts of this work are thus rooted in critical discourse and text analysis as a theoretical framework, as this approach emphasizes the strong connection of language and society.

The chapter begins with the description of the principles of feminist linguistics, followed by an outline of the different methods for gender expression in languages. It also illustrates generic masculinity with examples, explores theories behind gender stereotypes conveyed by language, and introduces the propositions of feminist language planning. I conclude the chapter with an overview of the previous research done on gender representation in the Estonian language and an overview of the attitudes toward gender issues within this language community, which are important for understanding the past, the present and the future of Estonian in relation to these matters.

### 2.1. The principles of feminist linguistics

The topic of gender is connected to language studies in feminist linguistics, which has been engaged with the question of how the positions of genders, gender inequalities as well as power relations between genders are reflected in language. This is in accordance with the principles of critical discourse analysis (see Fairclough 2001; Halliday 2007), where language is seen as strongly related to social practices. Therefore, language participates in constructing and reflecting social reality, as well as in conceptualizing gender. The earliest observations concerning gender in language can be traced back to the 1920s with Otto Jespersen's observations on women's language (1990[1922]). However, Jespersen's work lacks empirical evidence and instead reflects the norms and stereotypes of his time. In his writing, Jespersen discusses women's language use, focusing on its many aspects – the use of vocabulary, adverbs, phonetics, grammar, choice of words and many more. While his work was perhaps one of the first contributions to this field of linguistics, it is considered outdated and controversial by today's standards. It is the studies from the 1970s and 1980s, such as Robin Lakoff's research (1973) on the role of women in society as conveyed through language,

that are generally considered the core texts of feminist linguistics. Lakoff discussed the marginalization of women through language, arguing that language is used as a tool for maintaining and reinforcing it. She suggested that women experience gender discrimination, which is reflected in how they speak, how they are taught or expected to speak, and how they are spoken about. Lakoff's work has also been later criticized by other feminist linguists, due to her analysis relying on personal observations and intuition (Cameron 1992). However, it is still widely referenced today as an important cornerstone of theories on unbalanced power relations and patriarchal social order illustrated by language, contributing to the emergence of dominance theories of gender and language (West & Zimmerman 1987; Talbot 1998: 131).

Today, feminist linguistics examines both the use and production of language, in written and spoken forms, drawing on the hypothesis that linguistic practices reinforce the existing gender arrangements, and that language is an important tool for reflecting and recreating gender-related societal norms (see McConnell-Ginet 1988; Cameron 1992; Mills & Mullany 2011; Litosseliti 2013; Bucholtz 2014). At the same time, feminist linguistics is not only concerned with how language reveals beliefs about gender in a society, but also how these beliefs are communicated and how individuals express their own gender identities through language. According to Judith Butler's (1990) theory of the performativity of gender, gender is a series of actions and behaviors that individuals continuously perform, and these performances are based on culturally constructed norms. In other words, gender is something that people *do*, rather than something they *are*. It is often described as a fluid, shifting and changing phenomenon (Kelan 2010: 175), with language being an essential resource for performing gender on a daily basis. People actively construct their own gender, with language defining and delimiting it (Talbot 1998: 157). Therefore, language not only reveals how ideas about gender circulate in a society, providing means to shape perceptions about it, but also how it is used to manifest gender identities and performances of these identities (Baker 2008: 15). Since gender is often regarded as a natural occurrence – something taken “for granted as an obvious and invariable part of our identity” (Talbot 1998: 150), with beliefs about it being seen as the “obvious truth” (Eckert & McConnell-Ginet 2003: 9) – feminist linguistics approaches language with a critical focus, examining the conditions underlying a linguistic phenomena and questioning the normalized assumptions on which they are based (Cameron 1990: 2).

In 1980, Dale Spender wrote about ‘man made language,’ highlighting how men's dominance in power positions has granted them control over language production, resulting in a language that embodies a narrow, patriarchal worldview. Feminist linguistics views language as a conduit for this patriarchal perspective (Spender 1990). It is based on the assumption that language mirrors the world through a male standpoint, articulates it as experienced by men, and thus conforms to the stereotypical notions about genders and their unequal positions in the society. That is partly because men have historically dominated in language planning and regulation, such as creating dictionaries and writing grammars,

while women have contributed to language development through parenting roles (Cameron 1992: 4; Pauwels 1999; Pauwels 2003). As a result, language treats men as the norm, often including gender bias and being discriminatory towards women and other gender groups with lesser power. To denote such language, the term **sexist language** has been coined. Two forms of linguistic sexism can be distinguished: overt and covert sexism – in other words, the sexism inside as well as outside the language system. The first form denotes language that is inherently sexist, referring to men as the representatives of the norm, the “universal human being” (Koivunen & Liljeström 2003: 11), with other groups being inferior to them (Mills 2008: 11). An example of this is generic masculinity – such as using the word ‘man’ to denote ‘mankind’. The other form, i.e., covert sexism, indicates a situation where language becomes discriminatory in use, when a seemingly neutral expression is used in a sexist and discriminatory context (Thomas & Wareing 1999: 67–80). While overt sexism is thus illustrated by the use of explicitly gender-biased linguistic markers, covert sexism is more implicit, identified through surrounding context and often hidden behind humor and irony (Mills 2008: 34). Covert sexism can be exemplified with the practice of naming women with diminutives (e.g., *Kust sa lilleke siia said?* ‘How did you get here, little flower?’<sup>4</sup>) or by ironically referring to women with expressions of mock politeness (e.g., ‘dear’) (Wolfson & Manes 1980; Cameron 2024: 120; Kaukonen et al. 2024c). Both of these forms are represented in the Estonian language and are illustrated in the publications of this dissertation – examples of overt sexism include gender-marked words used in sports news, which I explain in P1, and occupational titles with generic gender suffixes, brought up in P4. The case study of the word *esinaine* (‘chairwoman’), which I describe in P2, represents both forms of sexism: it is overtly sexist, since it includes overt gender marker and is derived from a masculine word ‘chairman’. Furthermore, feminine forms can be discriminatory due to the less prestigious status they are often perceived as representing (Merkel 2013). However, the usage recommendation imposed on this term by language planners reflects covert sexism, as it assigns the word to a context of use that reinforces outdated gender roles. The use of compound words ending with *tädi* (‘aunt’) and *onu* (‘uncle’), described in P3, illustrates covert sexism, as this reveals how terms associated with children’s usage can also help to create negative association and convey a broader set of possible intentions when used by adults. In addition, it is important to note that sexist language can also be discriminatory towards men. For instance, in media portrayals, men are often described as embodying a specific type of masculinity, adhering to highly stereotypical attributes regarding appearance and characteristics (for example, Ricciardelli et al. 2010; Marling et al. 2024). Additionally, men are expected to convey authority and dominance through their language use (Kiesling 2007).

While language can be a means of gender discrimination, according to feminist linguistics, it can also convey **gender stereotypes**. Stereotypes differ from sexism

---

<sup>4</sup> This question was addressed to the former Estonian prime minister, Kaja Kallas, by a male member of parliament during a parliamentary question hour in 2021.

(both overt and covert), in that while they have damaging outcomes for individuals, they are not necessarily discriminatory or dangerous. Instead, they represent assumptions and generalizations that stem from categorizing, processing and interpreting reality (Lillhannus 2002: 5), which is a natural part of human behavior. Numerous studies reveal that stereotypical thinking develops from an early age (see Liben et al. 2002; Vervecken et al. 2013; Canessa-Pollard et al. 2022). In addition, stereotypes have been found to subconsciously influence behavior. For instance, Jung et al. (2014) discovered that the consequences of gendered naming of hurricanes can be fatal, since people perceive female-named hurricanes as less risky. Stereotypes are also slow to change, particularly those concerning men and masculinity, as men occupy fewer traditionally female-associated roles, while women have entered many traditionally male fields (Croft et al. 2015; Haines et al. 2016; Eagly et al. 2020). Stereotypes can be both descriptive (that is, describing individuals of different gender) and prescriptive (directing behavior by indicating and dictating how individuals should act) (Burgess & Borgida 1999; Heilman 2001). Stereotypes can be reflected in language in multiple ways. For example, vocabulary may reflect which occupations are traditionally more often associated with women or men. Gender stereotypes about roles and occupations reflected in language are described more thoroughly in Chapter 2.2.2 of this dissertation. Linguistic stereotypes in this dissertation are illustrated by P1–P3, which revealed stereotypes of female and male characteristics, roles and occupations expressed by the vocabulary, and by P4, which identifies signs of categorization of occupations into male and female categories by language users.

To sum up, as Litosseliti (2013: 9) has put it, the central concept of feminist approach to language involves “the shift from the view that we use language in certain ways because of who we are, to the view that who we are is partly because of the way we use language.” This shift underscores that language use is not merely an expression of identities, but our linguistic choices also play a role in constructing those identities as well as our interpretations of them. Thus, gender, while being described by language, is also formed and reinforced through it.

## 2.2. Gender in languages

Since gender is a fundamental social category, it is represented in every language in the world, in one way or another. Thus, languages can be divided into three groups based on gender expression. Firstly, in various languages, gender is incorporated into grammar. In such languages, nouns are classified into feminine, masculine and, in some cases, neuter gender, and this classification influences the agreement of adjectives, pronouns, verbs, and other parts of speech. These languages have a number of gender classes that are determined by certain markers (Hellinger & Bußmann 2015). These are known as **grammatical gender languages**, and examples include German, French, and Russian. Generally, the meaning of a personal noun determines its gender and the grammatical gender

corresponds with the sex of the individual it refers to (i.e., the word ‘mother’ is grammatically feminine) (Corbett 1991; Stahlberg et al. 2007: 164). However, in most cases, there is no inherent connection between the concept expressed by a word and its grammatical gender, illustrating how gender functions as a grammatical category. Secondly, languages known as **natural gender languages**, such as English and Swedish, mark gender primarily in personal pronouns (such as ‘he’ or ‘she’). The majority of personal nouns are gender-neutral in such languages and, in addition to gendered pronouns, gender is expressed through linguistic units referring to gender lexically (i.e., words such as ‘man’, ‘woman’). Lastly, there are languages without grammatical gender categories or any gendered pronouns – namely, **genderless languages** such as Finnish, Turkish and also Estonian, where most nouns, all pronouns and grammatical structures can be used to refer to individuals of any gender. In these languages, gender is typically indicated through lexical means that carry a semantic reference to gender (Stahlberg et al. 2007: 164–166). Gender-related information can also be expressed through address terms, including kinship terms (e.g., ‘aunt’ and ‘uncle’, see Clyne et al. 2009; ‘mother’, see Fitch 1991) and honorifics (e.g., ‘Sir’, ‘Madam’, ‘Mr’, ‘Mrs’), as well as through idiomatic expressions (Hellinger & Bußmann 2015). In addition, there exists a distinction beyond the typological level for differentiating the categories of gender expression in languages: grammatical, lexical, referential and social gender. Grammatical gender refers to gender as a grammatical feature, while lexical gender refers to lexical units that carry the semantic property of gender (e.g., terms such as ‘mother’, ‘Mr’). Referential gender pertains to linguistic expressions that relate to the non-linguistic reality and identify referents as female, male or gender-indefinite (e.g., pronouns ‘he’, ‘she’, ‘they’). Social gender represents socially constructed concepts related to gender, such as the association of roles, character traits and appearances with being either male or female, which are conveyed through language. Social gender is linked to stereotypical assumptions and gender roles, by creating a division between feminine and masculine (Kramaræ & Treichler 1985; Hellinger & Bußmann 2015; Cao & Daumé III 2020).

Thus, while some languages integrate gender in their grammatical structures, others differentiate genders only through semantic components. Importantly, as Stahlberg et al. (2007: 163) note, linguistic structures expressing gender do not only distinguish males and females, but they are also shaped by cultural and social understandings, traditional beliefs and even myths about genders within the language community. For example, in Dyirbal, an Australian Aboriginal language, birds are linguistically classified as feminine since they are believed to be the spirits of dead women (Corbett 1991: 16; Aikhenvald 2016: 121). Such linguistic structures in turn help construct and reinforce beliefs about how men, women and their roles are viewed.

Furthermore, grammatical gender can influence people’s perceptions about reality. For instance, speakers of grammatical gender languages have been found to associate inanimate objects with biological males or females, depending on the grammatical gender of the named objects (Phillips & Boroditsky 2003; Boroditsky

et al. 2003). However, gender expression (or the lack thereof) in a language does not automatically correlate with gender equality or neutrality in the society in which this language is spoken (Aikhenvald 2016: 217). Although there are studies revealing that countries in which grammatically gendered languages are spoken tend to exhibit less gender equality than countries where natural gender or grammatically genderless languages are spoken (see Prewitt-Freilino et al. 2012), linguistic practices reflecting gender inequalities exist in every language, including those without grammatical gender.

### 2.2.1. Generic masculinity

Regarding gender expression in languages, feminist linguistics has brought attention to a phenomenon known as generic masculinity, as an example of overt sexism reflected in language. Examples of generic masculinity appear in many languages, regardless of their typological classification. For instance, in English, the words ‘man’ or ‘mankind’ are used to denote people in general. In addition, the masculine pronoun ‘he’ was traditionally automatically used when the gender of the referent was not evident in the context (Silveira 1980: 166). In French, the masculine plural forms carry a double function, referring to either only males or both males and females (Xiao et al. 2023: 83). Additionally, the French masculine pronouns *il* and *ils* (‘he’ and ‘they’) have been regarded by the French language council Académie Française as the default forms in French, even recommending their usage in contexts where women are the majority in the referred group (Tudisco 2021). In German, grammatically masculine personal nouns (such as *Mathematiker* ‘mathematicians’) and pronouns (such as *seinen* ‘his’) are used to refer to a mixed-gender group or when the gender is unknown (Stahlberg et al. 2007: 169; Misersky et al. 2019: 2). Grammatically genderless languages also have lexical masculine generics. For example, in Finnish, compounds, derivatives or phrases referring to men (such as *uskottu mies* = *uskottu* ‘confidant’ SG.NOM + *mies* ‘man’ SG.NOM, ‘trustee, executor, administrator’) are used in a generic manner (Engelberg 2016).

This linguistic practice of attributing generic meaning to masculine words reinforces the notion, according to feminist criticism, that men are treated as the standard, while women (and individuals of other gender identities) are regarded as the exception or deviation from the norm. This pattern is evident, for example, in words referring to women which are always derived from a masculine word. Koivunen and Liljeström (2003: 11) refer to this as the “paradox of femininity”, since women are almost always positioned as secondary to men. Consequently, portraying men as the prototype of humanity results in a sexist bias that equates people with male individuals, ignoring women and other genders and rendering them invisible. When they are made visible, they are seen as distinct and marked (Silveira 1980: 166; Cooper 1989: 17; Pauwels 2003). Sally McConnell-Ginet (2020) suggests that marking and markedness are tied to social expectations, with marked elements representing what is perceived as unexpected. McConnell-Ginet further explains that historically, women, as the subordinate gender, have more

often occupied roles that were considered atypical or outside of what was traditionally expected from them, making them the more frequently marked category. As a result, their gender is more often emphasized, even in contexts where it should be irrelevant (McConnell-Ginet 2020: 50–51).

### 2.2.2. Gender stereotypes in occupational titles and role nouns

There is a famous gender riddle about a boy and his father who are in a car accident which only the boy survives. After the boy is rushed to the hospital, the surgeon there says “I can’t operate, that boy is my son!” Studies using this riddle have shown that many people do not initially solve it by suggesting that the boy could have two fathers or that the surgeon could be a mother (see Belle et al. 2021). This riddle is an example of occupational gender stereotypes that language participates in evoking and producing (Lewis & Lupyan 2020). Occupational titles and role nouns have been found to indicate the likelihood of specific words referring to either women or men (Irmén & Roßberg 2004), based on which occupations and roles are traditionally more associated with women or men. As the riddle shows, when processing language units, gender stereotypes linked to certain occupational terms are triggered in the minds of language users, making them associate specific occupational titles with particular genders (Carreiras et al. 1996; Kennison & Trofe 2003; Oakhill et al. 2005). According to Niedzwiecki (1993), occupational titles encompass many factors, including historical, linguistic, social, cultural and psychological elements. Therefore, they refer to more than just the occupation; they also evoke aspects such as the clothes worn on the job, tools and equipment, working conditions, social positions, duties, salary scale, and the individual fulfilling the role (Niedzwiecki 1993: 47–48). Consequently, a word referring to an occupation activates a mental image of a person of a certain gender in the cognitive interpretation of a language user – for instance, the word ‘surgeon’ indicates a male individual in many people’s minds. While the riddle represents occupational stereotypes that may arise independently of language, it is ultimately through language that these stereotypes are communicated and thereby reinforced. Without language, such mental associations would not exist in the same way.

Occupational titles have been a significant focus of feminist linguistics, as this field has taken special interest in traditional divisions and unequal power relations between women and men. Occupations are one of the core areas of gender stereotypes (Deaux & Lewis 1984) and also an important arena of ongoing social changes, which is also reflected in language (Holmes et al. 2009: 192). Language mirrors the prevalence of men in many positions, and the tendency to label occupational titles with a gendered word shows a deep embeddedness of gender roles in our semantic space (Penelope 1988: 121). Previous research indicates that titles referring to men designate higher prestige occupations. For instance, compounds ending with *man* have been found to denote positions of influence and leadership, law enforcement, protection and transportation, while occupational titles ending with *woman* indicate service-related tasks (Baker 2014).

Although Baker's corpus is historical, the findings illuminate the persistent influence of these associations in contemporary language use. Burnett & Pozniak (2021) found that prestigious universities in Paris, which are less activist, use more masculine forms than less prestigious universities, suggesting that masculine forms are seen as more prestigious compared to forms that also overtly denote women. Conversely, derivational suffixes marking females have been seen as trivializing women's occupations (Holmes 2001: 117). In some languages, these suffixes have a negative connotation, often resulting in derogatory usage (see Doleschal & Schmid 2001), and they have been found to be associated with a lower status (see Merkel 2013). Schulz (1990 [1975]) further discusses the phenomenon known as the semantic derogation of women, where terms referring to women (such as 'lady') undergo pejoration, eventually coming to denote lower positions ('cleaning lady') or even becoming insults. This confirms that professions are distributed along gender-stereotypical lines, revealing the societal power structures and male dominance embedded in language.

Social role theory (Eagly 1987; Eagly et al. 2000; Eagly & Wood 2012; Koenig & Eagly 2014) indicates that occupational gender stereotypes stem from people's observations and experiences about the roles typically performed more often by women or men. Historically, the allocation of these roles has been influenced by traditional approaches to biological sex differences between men and women (Levanon & Grusky 2016) and traits attributed to each gender. Men are socially perceived to have agentic traits, i.e., to be physically strong, authoritative, assertive, dominant and task-oriented, whereas women are expected to have communal traits, including being caring, sensitive, emotional and more service-oriented, due to their nurturing abilities (Heilman 2001; Abele 2003; Heilman & Eagly 2008). Such beliefs about occupational gender roles are often prejudiced and biased, not necessarily representing reality or reflecting the actual gender distribution in different occupations. Although these stereotypes may arise from individual experiences and observations, as social role theory suggests, they are also shaped by societal norms and are reinforced by various sources, including the media (Garnham et al. 2015; Gygax et al. 2016). As a result, although stereotyping is natural to human behavior, it frequently leads to overgeneralizations and discrimination.

Various studies have revealed how stereotypical information influences language processing. For instance, stereotypical information has been found to be stored differently from other semantic knowledge in a mental lexicon, often overriding grammatical cues (Molinaro et al. 2016). For example, speakers of Russian were found to make gender associations based on accurate gender distinctions of certain professions, even when the linguistic units were gender-contradictory and the generic masculine forms denoted feminine occupations (Kapatinski 2006; Garnham & Yakovlev 2015). Pyykkönen et al. (2010) found in an eye-tracking and visual paradigm study that Finnish participants were more likely to look at a character of the corresponding gender when hearing a gender-stereotype occupational noun (e.g., 'chimney sweep'), even in story contexts where this information was not needed to build coherence. This indicates that people draw on

stereotypical information, even when it is not crucial for understanding the story. In several studies, language users have processed both stereotype-congruent as well as incongruent word pairs, such as role noun followed by an anaphoric gender reference (e.g., ‘plumber’ + ‘sister’ or ‘babysitter’ + ‘he’). It has been found that mismatching pairs take longer to process and result in lower rates of acceptability and appropriateness (see Carreiras et al. 1996; Oakhill et al. 2005; Hammond-Thrasher & Järvikivi 2023), suggesting that stereotypical information is decisive. While these studies illustrate the significant influence of stereotypical information in interpreting role nouns, other studies, in contrast, have revealed that linguistic information has a more dominant effect on the interpretation. For example, studies using similar design in grammatical gender languages such as French or German have revealed that stereotype-incongruent anaphoric continuations are rated as inappropriate when the grammatical gender of the role noun mismatches the gender of the anaphora. Specifically, it was observed that when reading a sentence with a grammatically masculine generic role noun, female-referring continuations are more difficult to process and are rated as inappropriate, even when the role noun denotes stereotypically feminine occupations. The interpretations in such studies thus relied primarily on the grammatical gender of a role noun (Gygax et al. 2008; Gygax & Gabriel 2008; Misersky et al. 2019). Gendered conclusions drawn from role nouns can also result from the cooperation of two aspects: both language users’ individual world knowledge, including information derived from stereotypical assumptions, and language (Gabriel & Gygax 2008; Gabriel et al. 2017).

In such studies, particularly grammatically masculine generic role nouns have been the focus. This confirms that the generic masculinity is not, in fact, generic, as language users perceive it as primarily referring to males. Further empirical studies (see Schneider & Hacker 1973; Stahlberg et al. 2001; Gabriel et al. 2008; Gygax et al. 2009; Redl et al. 2021) have also found evidence that grammatically masculine generic nouns or pronouns are primarily interpreted as indicating male referents and that they reduce the visibility of other genders. This supports the feminist linguistic hypothesis that employing generic masculinity in language marginalizes women and people with other gender identities. It suggests that terms referring to men are perceived as the standard, leading to men being viewed as the default individuals, and reveals an inherent gender bias. Due to this evidence, masculine or male generics have also been referred to as ‘false generics’ (Hellinger & Bußmann 2015).

In addition, since occupational titles are often encountered in job advertisements, several studies have examined the effect of gender markings in this particular context and shown that gender markings, particularly generic masculinity, make representatives of other gender identities feel ignored and affect their motivation in terms of recruitment. For instance, as early as in the 1970s, Bem and Bem (1973) found evidence that gender-biased job advertisements discourage women and men from applying for jobs that were explicitly associated with a gender different from their own. In addition, language that directly refers to men (i.e., gender-exclusive language) can negatively affect women’s sense of

belonging and motivation in the workplace (Stout & Dasgupta 2011) and women have been found to be perceived as less suitable for high-status positions if the position in question is advertised in a masculine form (Horvath & Sczesny 2016). Not only nouns carry this bias. For instance, a study found that when a job ad includes adjectives linked to masculinity and male stereotypes – such as ‘assertive’, ‘independent’, ‘competitive’ – it affects women’s appraisal of the job, making them perceive themselves as unsuitable for these positions (Gaucher et al. 2011). Therefore, gender inequality is maintained in the language of job advertisements, often in a rather covert manner at first glance.

### 2.2.3. Feminist language planning

Jean Aitchinson (2000: 145–146) suggests that an important sociolinguistic factor behind language change is a social need: language changes due to changes in the needs and in the worldview of its speakers. Drawing from feminist linguistic theory, a movement known as feminist language planning (or feminist language reform<sup>5</sup>) has emerged across cultures and speech communities. It aims to change the way language represents and treats individuals of different genders – more precisely, to challenge gender representation in languages, offer solutions to reduce gender inequalities and sexism in language, raise awareness of gender bias, make other gender identities, besides men, more visible in language, to achieve a more balanced representation of gender in language (Pauwels 1999; Pauwels 2003). To summarise, feminist language planning is, in the words of Anthony J. Liddicoat (2011: 1):

“...an active engagement with the ways in which language represents and reproduces gender. It is not specifically a [sic] concerned with the ways in which language presents women, although this is a major focus, but rather how language positions both males and females and how it enters into the social practices that gender people and their activities and ideas. It is centred on a desire for equality between men and women as an overall social objective and for language to both support and reflect such equality”.

While Liddicoat explicitly mentions men and women, it is important to note that the focus of feminist language planning is also the linguistic representation of other gender identities. Thus, according to Aitchinson, as society becomes more open and aware, there is also a need for language that is non-discriminatory and

---

<sup>5</sup> In this dissertation, I refer to feminist language planning, as exemplified by Pauwels (2003), a term that is derived from language planning and corpus planning. According to Haugen (1983), language planning refers to a response to a social problem, and corpus planning, as part of language planning (Kloss 1969), involves modifications to language norms and structures. The goal of language reform is not simply to change language itself, but rather to focus on “the viability of change as well as on developing proposals for change” (Pauwels 2003: 552).

considers the interests of various identities, not just those in positions of power. This is what feminist language planning has aimed to offer.

Feminists began debates about regulating sexist language in the 1970s and 1980s (see Cooper 1989; Pauwels 2003: 552), emphasizing the exclusive effect of masculine generics and other linguistic units that imply men are the norm. Gabriel et al. (2018) have identified two strategies for such linguistic activism: feminization and neutralization. While many terms explicitly refer to male gender, feminization involves generating equivalent terms for females or generating word pairs, such as in French, *les étudiants et étudiantes* ('male and female students'). Word pairs can also be constructed as split forms, such as *végétarien/ne* ('male/female vegetarian'). In contrast, neutralization refers to relying on gender neutral terms and expressions, following the "idea of abandoning the explicit mention of female or male gender" (Gabriel et al. 2018: 850). An example for this would be using neuter gender in grammatically gendered languages, using non-gendered forms or forms that refer to people in general (i.e., *la population migrante* 'migrating population', instead of *les migrants/les migrantes* 'male/female migrants') (Stahlberg et al. 2001; Gabriel et al. 2018). Neutralization often also includes the elimination of feminine derivational suffixes (Pauwels 2003: 556).

Attempts to implement gender-aware language development strategies are evident in various language communities. For instance, in English, as a natural gender language, neutralization has been a primary strategy. Words with generic masculinity have been replaced by neutral counterparts (e.g., 'chairperson', 'chair' < 'chairman'), feminized suffixes have been ungendered ('flight attendant' < 'stewardess'), and gender-inclusive pronouns ('they/them') are used (see Romaine 2001; Bradley et al. 2019). These pronouns often serve a dual function: they are employed both as a replacement of the generic 'he' and to refer to non-binary individuals. Additionally, neopronouns, such as 'ze' or 'xe' have emerged (see Hekanaho 2022). According to Renström (2025), these could be considered as non-binary pronouns, as they are not used for a twofold purpose, but are specifically created to refer to non-binary gender identities. Similarly, in Swedish, another natural gender language, the gender-inclusive pronoun *hen* has been introduced instead of *hon* 'she' and *han* 'he'. It also serves a double function, being used generically, when the gender of the referent is irrelevant, or when referring to someone who does not identify as male or female (Gustafsson Sendén et al. 2015).

In grammatical gender languages, split forms, word pairs, gender asterisks, and other similar innovative symbols have been introduced. For example, in Spanish, letters denoting males or females have been replaced by symbols such as @ and X (*l@s estudiant@s/lxs estudiant@s* < *les/los estudiant@s* 'students'), or word pairs such as *les/los estudiant@s* are used (Kaufmann & Bohner 2014). In German, pair forms (*Spielerinnen und Spieler* 'female and male players'), gender gaps (*Spieler\_innen* 'fe\_male players'), asterisks (*Spieler\*innen* 'fe\*male players') and other solutions have been introduced (Diewald & Steinhauer 2020, cited through Freidrich et al. 2024). Neutralized forms for stereotypically female or male terms (e.g., *Timmer* < *Timmerman* 'carpenter') or combinations of female

and male suffixes (*Psycholo(o)g/e* ‘male/female psychologist’) have been proposed in Dutch (Gerritsen 2002). In Russian, feminists have attempted to generate and use feminized forms of role nouns (Berkutova 2019), as these are used less frequently than masculine forms and are often considered pejorative. Additionally, the use of gender neutral relative pronouns, such as *которые* ‘who, that, which’, as an equivalent to singular ‘they’, has been documented in the register of LGBTQIA+ communities (Kirey-Sitnikova 2021). Furthermore, while feminist language planning has focused on male-centered language that is perhaps more commonly represented by the use of role nouns and pronouns, the absence of names to denote the experiences of women and other gender identities is also a concern of this linguistic activism (Spender 1990). Therefore, the creation and adoption of feminist neologisms, such as ‘manspreading’, ‘himpathy’, ‘transmisogyny’, ‘cissplain’ to name those experiences represent another important dimension of feminist language activism (see Foubert 2024).

Thus, methods of implementing non-biased language (through either feminization or neutralization) varies from language to language. Since languages differ in terms of structure and the culture in which they are used, there is no one-size-fits-all solution, and all strategies have their pros and cons. For instance, feminization strategies improve the visibility of women in professions and other roles, while neutralizing strategies reduce the salience of the gender category and can therefore be generally more inclusive (Gabriel et al. 2018). However, feminine forms can unintentionally emphasize gender, have a negative connotation or can be associated only with specific reference (Pauwels 2003: 557; Mills 2008: 60; Formanowicz et al. 2012), while neutral alternatives may require more effort and time to process and gain acceptance (Gabriel et al. 2018: 851). Furthermore, even gender-neutral forms of masculine generics can still contain male bias (Lassonde & O’Brien 2013).

Despite widespread discussions about gender-neutral language, including the European Parliament’s publication of guidelines on the topic (EP 2018), and support for language planning, it has also remained a source of opposition and controversy, with examples of even governmental institutions intervening (in France, see La ministre de l’Éducation nationale et de la Jeunesse 2021). Resistance to feminist language planning is influenced by the global anti-gender movements, which are linked to the ideologies of right-wing populist parties and conservative social forces that emphasize, for instance, the very notion of gender as a threat to national identity or traditional values (see Wodak 2015; Kuhar & Paternotte 2017, Graff & Korolczuk 2022). Blaubergs (1980) argues that opposition to reforming sexist language stems from various arguments. These include the perception that promoting language change is an imposition, the belief that a language’s system for expressing gender is equated with gender equality (i.e., if a language lacks grammatical gender, the society is automatically more equal), the idea that language change is too difficult and impractical, and concerns about violating the historical authenticity of the language. Generic use of nouns and pronouns has, as Mills (2008: 51) indicates, traditionally been seen as built into the grammatical structure and thus difficult to change or regulate. Cameron

(2012: 238) indicates that while linguists are often neutral about language change, other language communities and users see it as an indicator of larger social change, thereby expressing anxiety about it. However, studies have shown that using gender-neutral or inclusive language instead of masculine generics does lessen male bias and has potential to weaken stereotypes (Vervecken et al. 2013; Horvath et al. 2016; Xiao et al. 2023; Kim et al. 2023). As Coutinho et al. (2024) have argued, while there is resistance to feminist language planning in both society and linguistic communities, a linguistic approach to gender-fair language should propose different solutions, acknowledge and document experimentation, and aim to understand and analyze these developments, since language contributes to the demands of equal citizenship.

### 2.3. Gender in Estonian

In terms of typological classification, Estonian, just as its linguistic relative Finnish, is genderless. Therefore, gender in Estonian is only expressed lexically. Firstly, full lexemes (e.g. words such as *mees* ‘man’, *naine* ‘woman’, *tüdruk* ‘girl’, *poiss* ‘boy’) can be used to indicate gender. Secondly, gender is expressed through compounding – either by using a gender-specific initial form or prefix, which acts as an adjective (e.g. *naispresident* ‘female president’, *meesarst* ‘male doctor’), or base form or suffix (e.g. *turvamees* ‘security man’, *esinaine* ‘chairwoman’, *õpetajaproua* ‘madame teacher’, *kirjanikuhärra* ‘mister writer’). While *mees* ‘man’ and *naine* ‘woman’ are the most common and perhaps most neutral gendered words in compounds, other gendered terms can also appear in the base form of compounds, such as kin terms like *tädi* ‘aunt’, *onu* ‘uncle’, or colloquial terms like *tibi* ‘chick’ and *kutt* ‘guy’. Lastly, derivational suffixes referring to a female agent are also used (e.g. *kuninganna* ‘queen’, *näitlejanna* ‘actress’, *prantslanna* ‘a French woman’, *sõbratar* ‘female friend, girlfriend’, *direktriss* ‘female director’) (Kasik 2015: 243–245; Hasselblatt 2015). There is no distinctive derivative indicating a male agent, which indicates that the base form by default refers to a male person. There is also even a derived version of the gender neutral pronoun *tema*, namely *temake*, which incorporates a diminutive suffix and is used to mark the contrast between ‘he’ and ‘she’ (Ross 1996: 104). However, it is not used frequently in contemporary language<sup>6</sup>. The standard status of words referring to men is also confirmed by usage frequencies, such as those found in the dictionaries, which have shown that units referring to women are used more frequently, as unmarked forms of words more often refer to men (for example, Hasselblatt 2015: 134).

In Estonian, generic masculinity is mainly represented in compound words ending with masculine noun *-mees* ‘man’, such as *esimees* ‘chairman’, *kalamees* ‘fishman’, *kaupmees* ‘merchant’, *asjamees* ‘specialist’, *abimees* ‘a helper’, *igamees*

---

<sup>6</sup> In the Estonian National Corpus (ENC) 2023, *tema* has around 30 million occurrences, while *temake* around 3500.

‘everyman’ etc. Generic femininity also occurs, but only in two compound words – *ämmaemand* ‘midwife’ (literally ‘lady of the mother-in-law’) and *medõde* ‘nurse’ (lit. ‘medical sister’). Furthermore, generic masculinity can also appear in certain idioms, such as *nagu üks mees* and *kõik see mees* (literally translated as ‘to a man’ and ‘all this man’, referring to doing something unanimously or all together), and *ole meheks* (lit. ‘be a man’, used in expressions of solidarity to indicate a request). There are also a generically used expressions such as *puhas poiss* (lit. ‘clean boy’, referring to a person or an organization that is perceived innocent) (see Piits 2015: 85).

Such gender markings have been introduced to Estonian on the example of other languages. For instance, the female derivational suffix *-nna* was adopted in the 19th century on the example of German, to refer to a woman of the upper class, and the *-tar* suffix was borrowed from Finnish in the 20th century (Kask 1955; Kasik 2004: 137). In languages without grammatical gender, gender-specific forms that are introduced can contain social gender bias, and such forms often highlight inequality rather than reducing it (Hasselblatt 2015: 133–134).

### 2.3.1. Research on language and gender in Estonian linguistics

One of the first studies documenting Estonian gendered vocabulary was by Kerli Puna (2006, also previously Olt 2004). She examined the use of gender-specific nouns for people in dictionaries and texts, using several Estonian dictionaries, the Corpus of Estonian Literary language as well as a cultural newspaper as data. Her study found that occupational gender stereotypes reflected in language are also apparent in Estonian. For example, Puna observed that as for compound words marking a person through an activity (i.e., occupation, position or group membership), gender markings align with traditional gender roles, and most of the professions have been traditionally held by men, which is why the corresponding words also include a masculine gender marker (e.g., *kalamees* ‘fisherman’, *jahimees* ‘hunter’, *meremees* ‘seaman’). Due to this, *mees* ‘man’-suffixed compound words also cover more areas of activity, since there was no corresponding term to describe a woman working in the same field. Additionally, it is not only the neutral words *mees* man and *naine* woman that convey stereotypes, but other gender-marked words as well. For instance, Estonian kinship terms *tädi* ‘aunt’ and *onu* ‘uncle’ in occupational titles indicate occupations considered either masculine or feminine – such as *sepaonu* ‘blacksmith uncle’ and *koristajatädi* ‘cleaning aunt’, with *tädi*-compounds referring mostly to education- and child-care-related professions. In addition, such compounds also indicate prestige, since *tädi* often represents less respected professions (Puna 2006: 39–41, 58, 88). Puna also indicates that, according to vocabulary, women are more family-oriented and associated with caregiving and household chores, while men are more represented in public life, dealing with agriculture, fishing, business, the military, etc. (pp. 39, 42). Furthermore, vocabulary describing appearances and other characteristics is highly stereotypical – men are described in terms of their height and strength, while women are characterized by their beauty, haircuts, body type,

etc. For men, traits such as wisdom, pugnacity, drinking, and making jokes are noted, whereas women are associated with gossiping, vanity, stupidity, and viciousness (pp. 37–38).

Stereotypical descriptions of women and men have also been documented in more recent analysis of sentences used as examples in definitions in the Estonian standard dictionary “Õigekeelsussõnaraamat”, in the study by Lydia Raadik (2016). In the example sentences of “Õigekeelsussõnaraamat 2013”, women are portrayed as moody, emotional and cold (e.g., *Hüsteeriline naine* ‘Hysterical woman’, *Tujukas tüdruk* ‘Moody girl’), while men are depicted as unemotional, naughty and angry (*Seda meest ei kõiguta miski* ‘This man is not shaken by anything’, *Poiss on koerust täis* ‘The boy is full of mischief’). Women are associated with humility, whereas men are linked to authority (Raadik 2016: 36–37). In addition, women’s speech is associated with, for instance, accusations, whining (*Naine vingub mehe kallal* ‘A woman whines at a man’) and prattling, while men’s speech is associated with drinking alcohol (*Purjus meeste jora* ‘A gibberish of drunken men’) and being straightforward (pp. 47–48). In terms of occupations, traditional gender divisions emerge, with men being portrayed in a more diverse set of roles than women. Men are associated with business, politics, leadership, construction and seafaring, women with childcare, handicraft, wild-crafting and acting. Men are also more associated with violence, criminality and drinking alcohol (pp. 59–61, 66).

A corpus linguistic approach to the analysis of gender representation in Estonian has been used by Liisi Piits (2015) in her doctoral thesis. By analyzing collocations<sup>7</sup> that occurred with words indicating people, including women and men, she discovered that socially constructed gender differences and gender roles are reflected in collocational relationships. For instance, the third most frequent collocate for the word *poiss* ‘boy’ was *paha* ‘bad’, indicating that younger males are often associated with misbehavior, while the third most frequent for *tüdruk* ‘girl’ was *ilus* ‘pretty’ (Piits 2015: 83). Collocates for the word *mees* ‘man’ described a male person through race, strength and height, whereas collocates for the word *naine* ‘woman’ indicated attractiveness and nudity. Furthermore, only women were described through adjectives mentioning family status or fertility, such as *vallaline* ‘single’, *endine* ‘ex’, and *rased* ‘pregnant’. In contrast, words indicating social status, such as *tähtis* ‘important’, *rikas* ‘rich’, and *vaene* ‘poor’ appeared only among the most frequent collocates for *mees*. Although the collocate *töötav* ‘working’ appeared with both words indicating men and women, in half of the cases referring to women, it was used to contrast women working outside of the home with homemakers (pp. 88–90).

While not strictly representing the field of linguistics, Barbi Pilvre’s (2009) dissertation is also relevant here. Pilvre analyzed feature stories of male and female individuals in the weekly Estonian newspaper *Eesti Ekspress*’ column

---

<sup>7</sup> Collocations, i.e. the lexical relationship of two words that are characterized by frequent co-occurrence, refer to word pairs that commonly appear together in language (see Stubbs 2001; Sinclair 2004).

named *Persoon* ('A Person'). Focusing on media representation, she found that women become newsworthy in profile feature stories when they succeed in professions perceived as traditionally masculine. In such cases, women are often portrayed as something unusual, a deviation from the norm, with an emphasis placed on their feminine qualities, such as looks, characteristics, and family life (Piltre 2009: 55–56). Similarly, Raili Põldsaar (2001) and Kristi Malmberg (2005/2006) have analyzed gendered discourses in Estonian media, revealing, for example, the trivialization of women (Põldsaar 2001: 104) as well as the blurring of the gender roles in the portrayal of men compared to the media of the 1990s (Malmberg 2005/2006: 31–32). Several later studies focusing on media representations of men and women have found that the representation practices tend to reinforce stereotypes – for instance, women are more often portrayed through private life, appearance and fashion choices, and women working on traditionally masculine positions such as leadership are perceived as exceptional (see Ester et al. 2020; Kaukonen et al. 2024a). In summary, studies concerning gender in Estonian remain limited, and those mentioned range from examining gendered vocabulary to the representation of men and women in the media, highlighting how gender roles are represented in the usage patterns of this language.

### **2.3.2. Attitudes towards gender in Estonian language users and communities**

Since language is a social and public act, its users have opinions about it, engaging in debates about what they think language is and what it should be. These debates reflect broader societal ideologies (see Cameron 2012). This also applies to Estonian, where anti-gender arguments have emerged from different perspectives.

The absence of grammatical gender in Estonian language has, on the one hand, contributed to an understanding that Estonian language is not gender-biased (see Koik 2022; Raudsaar 2023). On the other hand, it has resulted in discussions about sexist language as well as feminist language planning being underreported and underdeveloped. Raili Marling (2010) explains that the Soviet rhetoric in Estonia created a facade of gender equality, behind which deep inequalities were in reality hidden. For instance, while women did have access to paid work, they were still treated differently from men and were also responsible for domestic labour. After the collapse of the USSR, feminism was therefore seen as something “alien, dangerous and unnecessary”, becoming “a scapegoat for various social ills and a means for distracting attention from deep-seated gendered inequalities in the society but, possibly, also other social concerns.” (Marling 2010: 8)

The influence of this discourse is still relevant today. Although no systematic research has been done on these attitudes in Estonia, it is evident both in the previous statements of the language planners as well as in the public discourse that this has fostered a belief that gender inequality – including linguistic – is not an issue of concern, but rather a pseudo-problem (Põhjälä 2013). In a usage

recommendation published by Estonian language planners (Mäearu 2008), it is stated that occupational titles in Estonian are to be used neutrally. Therefore, it was advised to employ terms of generic gender, such as *esimees* ‘chairman’ or *meditsiiniõde* ‘nurse’ when referring to a person of any gender. Additionally, the term *esinaine* ‘chairwoman’ is to be used when referring to a chair of a female-only organization, and *meditsiinivend*<sup>8</sup> ‘male nurse’ (lit. ‘medical brother’) is considered inappropriate and informal (Õigekeelsussõnaraamat, i.e., Dictionary of Standard Estonian 1999–2018; Mäearu 2008; AMSS, i.e., Advisory Dictionary of Document Language). The belief that Estonian gendered terms are not gender-specific has been mentioned by scholars (Pir 2010: 145) and language practitioners have emphasized the gender-neutral nature of Estonian in contrast to other European languages, hinting that attempts to neutralize Estonian language are an imitation of other Western societies (Koik 2022).

Little public attention has been drawn to the gender bias of the Estonian language. Opinion pieces have been published on Estonia’s feminist popular science platform Feministeerium (Kuusik 2015; Kaukonen 2022), as well as in various media outlets (Papp 2013; Pakosta 2021; Kaukonen 2023). Such pieces have faced backlash from the representatives of Estonia’s right-wing parties, who have claimed that activists and scholars spreading these ideas are “raging feminists” attempting to attack and cancel Estonian language (Uued Uudised 2023), and by raising discussions on gender-based usage patterns in language, these activists are merely seeking “evidence for their talking mania, inferiority complex and gender-based bickering” (Makarov 2023). Additionally, claims about the sexism of the Estonian language have been countered with the arguments that Estonian language cannot be inherently sexist due to its lack of grammatical gender (Raudsaar 2023).

Another factor underlying this opposition can be summarized with the observation of Graff & Korolczuk (2022: 15): “‘gender’ is a stretchy category that serves as a screen for collective fears about change, loss of national identity, excessive influence of the West and its cultural hegemony.” Estonia, as a post-Soviet society, is characterized by a complex interplay between a longstanding separation from Western ideologies and a deep connection to national identity (Erdocia 2022) on one side, and the influence of Nordic countries on the other. Since the beginning of 2000s, following the regaining of independence in 1991 and joining the European Union in 2004, issues of gender equality started to be addressed. For instance, the Gender Equality Act was adopted and several institutions dealing with gender issues were established. However, gender-equality related movements have been and continue to be seen as a threat to national identity as well as language. In addition, Estonian has been significantly shaped by the standard language ideology (see Lindström et al. 2023), which affects

---

<sup>8</sup> Although *meditsiinivend* or *medvend*, referring to male nurse, is not common and illustrates more informal language, there are still instances of these words being used – in the ENC 2023, *meditsiinivend* occurs 18 and *medvend* 219 times (while *meditsiiniõde* occurs over 7000 and *medõde* over 5700 times).

discussions on gender, as speakers are particularly wary of language changes, especially those seen as politically motivated. Consequently, criticism of the gender-discriminatory aspects of Estonian has faced resistance. As Cameron (2024: 86) indicates, attempts to change language have always provoked conflict because they are perceived as undertakings with a broader political purpose. Although the European Parliament's gender neutral language guidelines have been translated into Estonian, discussions on this topic remain largely incomplete within both Estonian society and linguistics. The works referenced here are just a few examples of this debate.

The previous discussion has demonstrated that language as a system functions without requiring much conscious effort from its users. This indicates that language cannot be forcefully directed (Tavast 2022; Hayakawa & Hayakawa 2022: 101–102 [1949]). In other words, any adjustments proposed have to be acceptable and comfortable for users in order to take root, and language itself is indifferent to any discussions about it. However, language does not exist without its speakers; if stereotypes reside in the minds of speakers, stereotypes also manifest in language. In this regard, as Cameron (2024: 93) mentions, such efforts have had an effect on usage and improved the visibility of various identities, but the proposals to change language are more likely to succeed in societies which already support the aims of language planning activism. Hence, such discussions are necessary, to uncover the social information behind certain linguistic patterns, recognize the problematic aspects of gender-biased and sexist language, introduce the benefits of a more inclusive language and ultimately move toward using it. Such awareness is a crucial step towards achieving greater gender equality and inclusiveness – a goal I aim to advance with this dissertation.

### 3. METHODS AND DATA

In this dissertation, two different methodological approaches were utilized to address the research questions: corpus linguistic methods and a quasi-experimental method. In P1–P3, I relied on the former, extracting and analyzing data from linguistic corpora as the main source. To provide a comprehensive analysis of the corpus data, I employed a combination of methodologies, which were mainly based on corpus linguistics, such as corpus-based frequency and wordlist analyses. Additionally, I incorporated elements of corpus-assisted discourse studies to examine the contextual usage of specific vocabulary (in P1–P3) and to explore the social information that the language tends to convey. In P4, a quasi-experimental study design using a Likert scale survey was employed. This method, similarly to experimental studies, aims to evaluate representations and perceptions of linguistic units as well as their processing, to analyze the relationship between different variables, but without randomly assigning participants into control and experimental groups (Mackey & Gass 2021: 269).

The analysis employs both quantitative and qualitative methods. The quantitative approach is used to identify frequencies (P1–P3) and statistical significance of the frequencies (P1, P4) as well as the influence of possible variables on the survey results (P4). The qualitative approach is used to analyze usage patterns (P2, P3) and contextualize these patterns within a broader ideological context in the corpus analysis. Frequency data and statistics validate certain linguistic tendencies, but they alone do not fully capture the more complex and subtle practices of gendering in language and society. Thus, the qualitative analysis of the context of use helps to uncover more hidden discursive features in the language (McEnery & Wilson 2001: 76–77; Baker & Levon 2015: 233; Ancarno 2020: 175; Taylor & Heritage 2024). Combining quantitative and qualitative methods and linking interpretations thereby increases the reliability of the data and provides a more nuanced picture of certain linguistic phenomena (Baker & Levon 2015: 233; Partington et al. 2013: 10–12; McEnery & Baker 2015: 2). All of the data visualization, analysis and statistical tests in this study have been conducted with R (R Core Team 2023).

#### 3.1. Corpus linguistic study

Corpus linguistics can be defined as “the investigation of linguistic research questions based on the complete and systematic analysis of the distribution of linguistic phenomena in a linguistic corpus” (Stefanowitsch 2020: 55). Corpus linguistic study relies on the examples of natural, authentic language (McEnery & Wilson 2001: 1), as represented in large collections of digitized texts, which are analyzed to uncover language characteristics through the typical patterns of association reflected in them (Biber et al. 1998: 4). Research questions often emerge from the intuitions of researchers and their observations about a language

phenomenon (Conrad 2002: 77). In corpus research, two main approaches are distinguished: corpus-based and corpus-driven studies (Tognini-Bonelli 2001: 65–66, 84–85). Corpus-based research uses corpora to address preformulated research questions or to explore hypotheses, while corpus-driven research allows the data itself to guide the discovery of patterns, providing empirical evidence without the influence of prior hypotheses or linguistic theories. My studies have been corpus-based, as I have approached the data with the hypothesis rooted in feminist linguistics, that language conveys gender stereotypes.

Since corpus linguistics is a form of text linguistics, where texts are believed to serve as reflections of interactions between writers and readers, they also reveal ideologies and social understandings communicated by the writers (Martin & Rose 2007: 1, Partington et al. 2013: 5). As such, corpus linguistics is often integrated with discourse analysis in a method known as corpus-assisted discourse studies (or CADS). This approach views texts as discourse that conveys social information, using software to identify language elements, their frequencies and patterns between them that might reflect this information. Thus, while critical discourse analysis explores how ideologies, social and cultural practices, power relations and values are embedded in language use (Wodak 1989; van Dijk 1993; Fairclough 2001), CADS adds the advantage of large-scale corpus data to support this critical analysis (McEnery & Baker 2015; Partington et al. 2013; Ancarno 2020). According to Conrad (2002: 78), corpus linguistics is “particularly helpful in providing “big picture” perspectives on discourse – determining patterns of language behavior across many texts, identifying typical and unusual choices by users, and describing the interactions among multiple variables”.

Gillings et al. (2023) note that while research incorporating CADS generally centers on social rather than linguistic questions, such studies can nevertheless be equally motivated by interest in social practices, power imbalances, and the language use that reflects and expresses these dynamics. As also Partington (2023) highlights in his book review, CADS therefore offers a variety of possible approaches, with its methodologies having a broad applicability. Its emphasis can also shift toward language research methodologies rather than social issues. That said, in this dissertation I use CADS to investigate the frequencies and frequency ratios of gender-marked words, while also analyzing usage patterns in the context, how they have evolved and what they tell us about gender roles in the society. Positioned along the “corpus-discourse spectrum” (Taylor & Heritage 2024) that is CADS, my work leans more toward the corpus side. However, as I am also interested in social questions, such as the role of Estonian language in constructing and perpetuating gender-related issues, I draw on aspects of the CADS approach within my research framework. Moreover, while the focus of this work is not on identifying discourses, which is why it is not perhaps the most conventional in the CADS tradition, the emphasis of CADS on analyzing representations to uncover how language constructs inequalities (Taylor & Heritage 2024) aligns closely with the objectives of my work, as I examine the representations of women and men in Estonian vocabulary and its reflection of gender inequalities. Such an approach has its limitations, which I will discuss in the Chapter 5.6, suggesting

incorporating a more discursive perspective in future research to better uncover the deeper societal implications of language.

### 3.1.1. Used corpora

The information about corpora used in publications P1–P3 is summarized in Table 1 below.

**Table 1.** Overview of corpora used in the dissertation.

Publication	Used corpora	Corpus sizes (in words)	Year <sup>9</sup>
P1	Delfi sports	436 582	2020
	ERR sports	401 880	2020
	Estonian Reference Corpus (ERC)	203 267 951	1990–2008
P2	Digital Archives of Estonian National Library	~648 800 850	1920–1979
	ERC	203 267 951	1990–2008
	web subcorpus 1 of Estonian National Corpus (ENC) 2021	~247 909 932	2013
	web subcorpus 2 of ENC 2021	~723 822 032	2021
P3	web subcorpus 2 of ENC 2021	~723 822 032	2021

In P1, I used three different corpora to examine gender-marked vocabulary: two corpora consisting solely of sports news from the Estonian web news portals Delfi and ERR as well as the Estonian Reference Corpus. The sports news corpora include texts published from January to March 2020. A total of 5032 articles was scraped from the web and compiled into a “self-made” corpus. The Estonian Reference Corpus contains texts from the years 1990–2008, specifically, journalistic texts from 1995–2008, fiction, parliamentary transcripts from 1995–2001, laws, scholarly texts (dissertations, journals etc.) and new media (forums, chat rooms etc.) (Koppel & Kallas 2022: 218).

To analyze whether the usage of the word *esinaine* (‘chairwoman’) has changed over the century in P2, the corpora included Estonian newspaper publications, such as *Edasi*, *Vaba Eesti Sõna*, *Postimees*, *TRÜ Ajaleht*, *Päewaleht*, *Kaja* and others, from six decades from the 1920s to the 1970s. I accessed the data through the digital archives of the Estonian National Library<sup>10</sup>, with the help of Tinitis (2020). Additionally, I used the Estonian Reference Corpus as well as two

<sup>9</sup> As for the web subcorpora of the Estonian National Corpus, this reflects the year the texts were collected, whereas for the rest, it refers to the year of publication of the texts

<sup>10</sup> <https://dea.digar.ee>

subcorpora of the Estonian National Corpus 2021 for comparison with the earlier years. The latter corpora include web texts from periodicals, forums, blogs, chat rooms, web pages, web commerce etc., scraped and compiled in 2013 and 2021, respectively (Koppel & Kallas 2022).

For P3 that involved examining frequencies as well as usage patterns of gendered vocabulary ending with lemmas *tädi* ('aunt') and *uncle* ('onu'), I also used the web corpus from 2021.

### 3.1.2. Data extraction and analysis

For data collection from the corpora, I used the SketchEngine software<sup>11</sup> (Kilgarriff et al. 2004; 2014). First, I employed the Wordlist function to generate frequency lists of gender-marked vocabulary for P1 and P3. This function makes it possible to find a list of words beginning, ending and containing specific characters. In P1, I queried words that begin with the lemmas *mees* 'male' and *nais* 'female' as well as end with the lemmas *mees* 'man' and *naine* 'woman', in both sports news corpora and the Estonian Reference Corpus. Since the Estonian Reference Corpus includes texts on various themes, I manually filtered the results to retain only occurrences related to sports contexts (i.e., the same words that appeared in the sports news corpora). For P3, I primarily searched for words ending with the lemmas *tädi* 'aunt' and *onu* 'uncle', but I also included *naine*, *mees*, *tüdruk* 'girl', and *poiss* 'boy'. For both publications, I exported the results along with their frequencies as Microsoft Excel files and then analyzed the data in greater detail. This involved reviewing the data to remove incorrect word combinations, such as typos (i.e., *häärastmees* = *häärasmees* 'gentleman'), word pairs (*turvamehed-naised* 'security men-women') and recurring or similar instances (*showmees* 'showman' includes *šõumees*, *sõu show-meess*, *shõumees* etc.). Since a large number of words appeared, manual review may have introduced some variation in the reported frequencies due to potential human error. The frequency analysis focused on both token and type frequencies. Token frequency refers to the absolute number of occurrences of gendered compound words in the corpus, while type frequency indicates the number of distinct gendered compounds (Stefanowitsch 2020: 311–313). In P1, statistical significance between the token frequencies was tested using the Chi-square test.

To understand what these words convey in both P1 and P3, I categorized the words based on the first part of the compound, which reflects a characteristics or type of a male or female person. In Estonian, compound words are formed by naming things, phenomena, activities, etc., through a characteristic connection, which can be direct, indirect, or based on similarity (EKG I: 407). Each compound word was thus placed into a semantic category based on the meaning of its first part, and ranked by frequency. In P3, I further examined words specifically

---

<sup>11</sup> <https://www.sketchengine.eu>

related to occupational titles by “zooming in” on this semantic group to analyze which occupations are represented in the data.

The qualitative approach of the work is represented by concordance analysis, which I employed to examine usages of selected words in P1–P3. Concordance is a function that displays the queried linguistic unit within its surrounding context. In P1, I reviewed a few sample sentences for each gender-marked word to understand the context in which these words were used. For P2–P3, sentences for a more systematic analysis of usage were collected using SketchEngine’s “Get a random sample” function. Table 2 below shows the total number of sample sentences downloaded and analyzed from different corpora.

**Table 2.** Data for analyzing usage in P1 and P3.

Publication	Key word in context	Corpora	Number of occurrences in total	Number on analyzed sentences
P2	<i>esinaine</i> ‘chairwoman’	a) 1920–1949 b) 1950–1979 c) ERC, web 2013, 2021	a) 8986 b) 53 c) 7800	a) 300 b) 53 c) 300
P3	a) <i>kokatädi</i> ‘cook aunt’, <i>söögitädi</i> ‘dining room aunt’, <i>arstitädi</i> ‘doctor aunt’, <i>raamatukogutädi</i> ‘library aunt’, <i>koristajatädi</i> ‘cleaning aunt’  b) <i>politseionu</i> ‘police uncle’, <i>arstionu</i> ‘doctor uncle’, <i>korraldajaonu</i> ‘organizer uncle’, <i>miilitsaonu</i> ‘militia uncle’, <i>valvurionu</i> ‘guard uncle’	web 2021	a) 1196 b) 398	a) 500 b) 293

In P2, I analyzed sample sentences by decade. I queried the more recent corpora’s (1990–2008, 2013 and 2021) concordance with the lemma *esinaine* ‘chairwoman’, creating a random sample of 100 sentences from each decade and exporting the results. The data from the years 1920–1979 consisted of pre-loaded concordance lines in Excel files that were not available on SketchEngine; thus, random samples from this corpus were generated manually, with the help of Excel’s random sample function. Less material was available in the corpus of the years 1950–1979 due to the limited number of digitized and accessible publications, and fewer articles published during that period. Therefore, I analyzed all usage examples from this time.

In P3, I examined random sample sentences of the 5 most frequent *tädi*- and *onu*-ending occupational words that are listed in Table 2, selecting 100 sentences with each word. Some of the selected occupational titles ending with *onu* had fewer than 100 occurrences in the corpus. The qualitative study in P3 also employed thematic analysis (see Braun & Clarke 2006; Clarke & Braun 2017) to identify and analyze themes of different usage patterns. In both P2 and P3, each concordance line was manually annotated in an Excel file regarding the usage it represented and then analyzed to determine the social information underlying such usage patterns.

Although the number of sentences analyzed is selected systematically, using random sampling to extract data from diverse sources and contexts within the corpora, some unevenness in the sufficiency must be acknowledged. While the distribution of samples analyzed is mostly even across subcategories, this was not always achievable. For example, *esinaine* has significantly fewer occurrences in 1950–1979 and compounds ending with *onu* have fewer occurrences in the web corpus of 2021, due to the reasons mentioned above. While the scarcity of *onu*-compounds is an interesting finding in itself, it may also indicate a limitation, as there were fewer examples analyzed than for *tädi*. However, while representativeness is a key feature of a corpus, it is often unavoidable that data on some linguistic phenomena may be incomplete, as the datasets cannot capture every possible language variant or sublanguage. In these cases, generalizations can be drawn from incomplete data and be refined as more data becomes available, since the incompleteness is simply often a common characteristic of a linguistic corpus (Stefanowitsch 2020: 5, 7).

I am aware that in a qualitative approach, such as the analysis of usage, interpretations of the observed usage patterns often rely on the researcher's intuition and introspection. However, while it is not always possible to completely remove introspection, it can be limited to help reduce subjectivity (Klavan et al. 2013: 18; Mautner 2009: 45; Baker 2006: 18). This can be done with the help of the conclusions drawn from the surrounding context, the researcher's native knowledge of the language as well as the large datasets of language units that a corpus provides (Stefanowitsch 2020: 15).

### **3.2. Quasi-experimental study: Likert scale survey**

To investigate gender perception of occupational titles in Estonian and Russian in P4, a study design employing a Likert scale survey was used. The Likert scale is an ordinal scale, i.e., a type of qualitative scale typically presented in the questionnaire format. It allows participants to rank their responses (e.g., from 'strongly agree' to 'strongly disagree'). Values of the scale can be ordered, though the intervals between values cannot be precisely measured, as they are not necessarily equal (Sullivan & Artino Jr 2013; Gillioz & Zufferey 2020: 33). Originally developed for studying attitudes (Likert 1932), the Likert scale is also suitable for studying perceptions of gender. In previous studies examining gender perception in role names, similar response scales have been used (see Gabriel et al. 2008;

Misersky et al. 2013; Tibblin et al. 2023; Gabriel et al. 2023). Although they have slightly differed from the classic 5- or 7-point Likert scale, incorporating more points and employing numerical values by asking participants to indicate the percentage (from 0–100) of women or men they believe are represented by specific role nouns. Such questionnaires, including the one used in my dissertation, illustrate offline measures for studying language comprehension. Hence, the focus of the analysis is on the final interpretations resulting from the language comprehension process (Gillioz & Zufferey 2020: 17).

In P4, two 5-point Likert scale surveys were employed, one in Estonian with 36 sentences featuring 36 occupational titles, and the other in Russian with 34 sentences featuring 34 occupational titles. The selected titles represented different morphological structures of gender expression (i.e., gender neutral vs gender-marked occupational titles in Estonian and grammatically masculine vs feminine titles in Russian) as well as different types of occupations based on real life gender ratios from Estonian and Russian statistics sites (i.e., male-dominated, female-dominated and balanced-gender occupations). The sentences were assigned to participants in a random order, and all participants rated all 36 words. Therefore, I classify this study as quasi-experimental, as no separate groups (control vs. experimental) were used, and the stimuli were not divided across participants according to different conditions. In the instructions, participants were informed that the occupational titles they would see in the survey referred to hypothetical recruitment contexts, as this context best illustrates the emergence of stereotypes and the impact of bias indicated by language. The task was to imagine that different companies were hiring for these positions, and to indicate whether men or women are preferred for each role, choosing from the options *ainult mehi* ‘only men’, *pigem mehi* ‘rather men’, *pole vahet* ‘doesn’t matter’, *pigem naisi* ‘rather women’ and *ainult naisi* ‘only women’.

Using such scale-oriented surveys requires consideration of central tendency bias, where participants tend to place their responses closer to the midpoint to avoid giving extreme responses at the end of a scale (Malone et al. 2014; Kusmaryono et al. 2022). This bias may also occur when participants lack strong opinions or interest (Krosnick et al. 2002; Nadler et al. 2015), or when they aim to provide socially acceptable answers (Garland 1991). However, it has been suggested that the central tendency bias is an inherent outcome of how participants interpret such scales (Douven 2018).

### 3.2.1. Data collection and analysis

The surveys were hosted on the LimeSurvey<sup>12</sup> platform and were available for completion from October to December 2023. Participants for the Estonian survey (a sample of 581 native speakers in total) were recruited through social media platforms such as Facebook, Instagram, and Reddit. The Russian survey

---

<sup>12</sup> <https://survey.ut.ee/>

(326 participants in total) was shared on Facebook, Vkontakte and Telegram. Prior to public distribution, pilot tests were conducted with 7 Estonian participants and 20 Russian participants. Based on the feedback from these pilot tests, adjustments were made before the surveys were shared publicly. Incomplete responses, as well as responses from participants under the age of 18, were excluded from the analysis. In addition to the survey questions, participants were asked about their gender, age, place of residence, and education, to assess whether these variables had an impact on their responses.

Statistical data analyses were conducted using R. In the analyses, frequencies of responses were firstly reported, with responses categorized into three groups: male-biased (combining ‘only men’ and ‘rather men’), neutral and female-biased (combining ‘only women’ and ‘rather women’). A Chi-square test was used to find statistical differences between the three groups. A principal component analysis was performed on all of the 581 Estonian and 326 Russian responses to the 36 and 34 occupational titles, respectively, using the “fviz\_pca” function in the FactoMineR and FactoExtra packages, to explore whether stereotypical information or language influenced the distribution of responses. To further identify potential predictors, such as participants’ demographic data, logistic regression models were generated using the “glm” function. Two binomial models were run for both languages, comparing neutral responses (set as the reference category) with male- and female-biased responses.

The survey data differs from the corpus data, as it reflects the assessments of language units guided by the survey questions, rather than the spontaneous and natural language use that could be expected from the corpora. Hence, the differences in how the data was collected and the different contexts they represent make it challenging to equate these sources. Furthermore, there is no demographic information available of authors of the corpus texts, and it cannot be precisely determined when these texts are written. Therefore, it is not possible to draw direct comparisons between the two datasets. However, both sources revealed many instances of negative attitudes toward the subject of gender and gender equality. It could also be argued that since participants were primarily recruited from social media platforms and since the web corpora include texts from blogs, forums, and commentaries, both datasets reflect language use by individuals in similar digital contexts.

## 4. OVERVIEW OF THE PUBLICATIONS

**P1. Kaukonen, Elisabeth. 2022. Sooliselt markeeritud sõnad eesti spordi-uudistes. [Gender marked vocabulary on the example of Estonian sports news]. *Keel ja Kirjandus* 6, 526–545**

In the first publication, I analyzed gender marked vocabulary in the sports news of web portals Delfi and ERR. My aim was to first examine the frequencies of gender marked compounds and secondly to find out which gender-specific compounds are used in the sports context. I further examined sports-related gendered vocabulary in the Estonian Reference Corpus (1990–2008), in order to compare the vocabulary that emerged from texts published in 2020 with earlier years. I relied on the feminist linguistics principle that language use reflects various gender stereotypes, which can be reinforced through lexical gender markings. The topic of sports was selected, as this field is traditionally strongly associated with masculinity and still exhibits considerable segregation and gender inequality (Schneider 2000; Mawson 2006; Zamfir 2024).

The analysis of frequencies revealed that in vocabulary ending with gender marking, words referring to women (i.e., compounds ending with *naine* ‘woman’) accounted for significantly fewer occurrences than those referring to men (i.e., compounds ending with *mees*). In contrast, for compounds beginning with gender specific initial forms *mees* ‘male’ and *nais* ‘female’, the token frequencies were more balanced. I concluded that this confirms the generic nature of *mees*-ending compounds, as they are more often used to denote people, i.e., athletes in general, indicating deeper lexicalization as opposed to compounds ending with a feminine gender marking. Additionally, since compound words beginning with gender marking generally refer to specific roles (i.e., *naisujuja* ‘female swimmer’), the higher frequencies of *nais*-beginning compounds suggests that the basic meaning of such words without a gender marking (*ujuja* ‘swimmer’) typically refers to a male individual. This illustrates how women are traditionally seen as deviations from the norm in sports. It also indicates that these gender markings exhibit different nuances in meaning and register. Gender marking in the base form of a compound word typically represents a more colloquial expression, whereas gender emphasis in the initial form tends to be more formal. In the latter case, gender does not define the main meaning of the word, but instead serves as an attribute that is used to describe a role noun. Furthermore, a comparison with texts from an earlier period revealed a slight increase in the use of gender marked words.

While these results could simply reflect the sports context – stemming from the fact that sports journalists are mostly male, the authors may use such words for stylistic purposes (e.g., to vary synonyms) and because men’s sports potentially receive more media coverage – they nevertheless characterize gender roles and positions of men and women in society, highlighting how different language use patterns convey these dynamics. Additionally, sports news remains an important part of public discourse, shaping reality through its language use.

**P2. Kaukonen, Elisabeth. 2023. Kes on esinaine? [Who is Chairwoman?] *Keel ja Kirjandus* 3, 328–336**

The second publication was motivated by a 2008 recommendation of Estonian language planners, according to which the word *esinaine* ‘chairwoman’ should be used only when referring to the director of a female-only organization (such as *naiskoori esinaine* ‘the chairwoman of the women’s choir’). In all other contexts, using the word *esimees* ‘chairman’ was advised, as it was considered gender neutral in Estonian (Mäearu 2008). In this publication, I analyzed data from corpora spanning different periods of the past century (1920s–1940s, 1950s–1970s, 1990–2008, 2013 and 2021) to determine whether *esinaine* occurs in its recommended meaning only in earlier usage, or if this usage persists in modern language. I hypothesized that this recommendation is no longer relevant in contemporary language use.

I examined concordance lines from the texts of each period, using a random sample method, and coded each sentence based on whether *esinaine* refers to a mixed-gender or female-only organization. The results showed that the recommended usage was dominant in the 1920s–1940s, with over 90% of the analyzed examples indicating that *esinaine* referred to a director of an all-female organization. However, from the 1990s onward, *esinaine* has increasingly been used to refer to the leader of mixed-gender organizations. Therefore, the analysis revealed that the recommendation is not relevant in modern language, as *esinaine* is now frequently used to refer to the director of any organization. This shift is appropriate, reflecting the fact that women today can occupy leadership roles, which was not common 100 years ago. In 2022, the Estonian Language Institute updated its recommendation in their combined web-based dictionary accordingly.

**P3. Kaukonen, Elisabeth 2023. Cleaning aunts and police uncles in action. Unveiling gender dynamics in Estonian compound words. *Eesti ja Soome-Ugri Keeleteaduse Ajakiri (Journal of Estonian and Finno-Ugric Linguistics)* 14(3), 137–171.**

This study explores Estonian compound words ending with the lemmas *tädi* ‘aunt’ and *onu* ‘uncle’. The aim was, first, to determine the frequencies and semantic categories of such compounds and, second, to analyze how occupational titles that incorporate these gender-specific words are used in contexts. The data consists of the web subcorpus 2021 of the Estonian National Corpus of the same year. The methods employed were corpus-assisted discourse studies, which combined corpus linguistic and discourse analysis techniques to examine frequencies and interpret concordance lines, and thematic analysis, to identify and analyze themes of usage patterns.

*Tädi* and *onu* in Estonian represent fictive kinship – that is, while they primarily denote kin relations, they are also used when no kin relationship or sometimes even any acquaintance exists between the language user and the referent

(see Gu 1990; Fitch 1991). In some cultures, including Estonian, *tädi* and *onu* are used in children's language, often to address unfamiliar people older than the speaker (for Swedish, see Clyne et al. 2009; for Tatar, see Kiss 2022; for Estonian, Puna 2006: 58). Generally, this usage derives from the speakers' aim to express politeness, affection or respect. However, it can also convey a more derogatory nuance.

The results showed that compound words ending with *tädi* have slightly higher frequencies than those ending with *onu*. Furthermore, the occupational titles incorporating such words reveal how such vocabulary expresses traditional gender roles and real-life employment-based gender segregation in Estonian society (Statistikaamet 2022), with women dominating in the educational sphere and social welfare, while men are predominant in construction. According to the findings, *tädi* also appears more frequently in occupations that are lower-paid and tend to be associated with lower prestige (Schulz 1990 [1975]). The analysis of usages revealed that while *tädi* and *onu* often appear in children's language to express familiarity and warmth, they can also appear humorously or carry negative connotations, sometimes serving as insults. These findings indicate that gender differentiation also becomes evident in languages such as Estonian, which is typologically genderless and thus considered relatively gender neutral.

#### **P4. Kaukonen, Elisabeth, Polina Oskolskaia, Liina Lindström & Raili Marling 2025. Gender, language and labour: Gender perception of Estonian and Russian occupational titles. *Frontiers in Communication* 9**

Research on occupational gender stereotypes in language has suggested that gender bias in language users' perception arises from factors such as beliefs about roles traditionally associated with either men or women and linguistic cues. Building on these insights, this study examines gender perception among speakers of Estonian and Russian. The aim was to determine whether occupational gender stereotypes in these languages are influenced by social knowledge and/or by language. Since similar research questions have been studied more extensively in grammatically gender languages, this study emphasizes understanding gender perception in a grammatically genderless and understudied language like Estonian, comparing it with a linguistically distinct, non-Indo-European language such as Russian. We also hypothesized that masculine generics would be perceived as referring specifically to male individuals, despite their intended neutral usage.

Two separate online Likert scale questionnaires – one in Estonian and one in Russian – were conducted to answer the research questions. Data from 581 Estonian and 326 Russian participants was collected and analyzed statistically. The results revealed that while most occupations were perceived as gender-neutral in both languages, biased perceptions still exist. These biases appear to stem from occupational gender stereotypes, particularly in the case of Russian. However, the findings also suggest that language influences gender perceptions, and that generically used gendered titles are perceived as more male-biased than neutral

forms in Estonian. We concluded that although Estonian is genderless and thus assumed to be more gender neutral than gendered languages, its vocabulary, which includes explicit gender markers, can evoke as much bias as grammatically gendered Russian. In Russian, grammatical gender represents a more deeply entrenched linguistic structure. As a result, in the case of generic masculinity, linguistic gender information may be less salient when interpreting the social gender of occupational titles in Russian. Nevertheless, Estonian language users are less often guided by linguistic gender information due to the absence of grammatical gender. This absence allows for more neutral interpretations than in Russian, as evidenced by the more neutral responses to morphologically gender-neutral occupational titles in Estonian compared to Russian. In sum, our results demonstrated that exploring gender perceptions in a more diverse range of languages is necessary, as it deepens the understanding of how gender functions across societies.

## 5. RESULTS

The results of this dissertation are presented in the order of the research questions formulated in the introduction, with each subchapter of the results section addressing one research question. The research questions were:

1. What gender-marked nouns referring to people occur in Estonian and what is the ratio of words referring to women and men?
2. Have the frequencies and usage patterns of gender-marked nouns changed over time? To what extent do shifts in usage reflect evolving societal gender norms and align with the principles of feminist language planning?
3. Which semantic categories are represented by Estonian gender-marked nouns referring to people?
4. Which occupations are reflected in the gendered vocabulary marking occupations?
5. How do language users perceive occupational titles in Estonian and is this perception influenced more by language or stereotypes? How does this differ from the perception of occupational titles in Russian?
6. Does the usage of gendered words reveal the positions of women and men in the society and how does Estonian gender-marked vocabulary convey gender stereotypes and biased representations of men and women in general?

This section begins with a general overview of the results, covering frequencies, changes in the vocabulary, semantic categories, types of occupations, and the perception of occupational titles. These aspects are addressed by answering research questions 1–5. In Subchapter 5.5, I synthesize the answers from all previous subsections and research questions to draw conclusions about the implications of the results and what they reveal about gender bias in the Estonian language. In that subsection, I specifically address research question 6 by integrating findings from all publications. Finally, in Subchapter 5.6, I discuss the limitations of the study and suggest directions for future research on this topic.

### 5.1. Frequencies of gender-marked words

I first present the frequencies of the analyzed gender-marked words on the basis of three different corpora: sports news (Delfi and ERR) and the Estonian Reference Corpus, examined in P1, as well as 2021 web corpus, analyzed in P3. Type and token frequencies of the following word groups were examined:

1. compound words ending with *mees* ‘man’ and *naine* ‘woman’ (P1, P3);
2. compound words beginning with *mees* ‘male’ and *nais* ‘female’ (P1);

3. compound words ending with *poiss* ‘boy’ and *tüdruk* ‘girl’ (P3);
4. compound words ending with *tädi* ‘aunt’ and *onu* ‘uncle’ (P3);

The frequencies of words ending with *mees* and *naine* cover two contexts: sports-related words in P1, and general use in P3. Although the words in question do not cover the entire gender-marked lexis of the Estonian language, they represent some of the most significant vocabulary groups.

The frequencies are summarized separately, regarding the context they represent. Firstly, the frequencies of sports-related vocabulary are summarized in Table 3 below.

**Table 3.** Frequencies of sports-related gender-marked vocabulary.

Word group	Sports news (Delfi and ERR) 2020				Estonian Reference Corpus 1990–2008			
	Token freq.	Type freq.	Ratio	Relative freq. (per 10000 words)	Token freq.	Type freq.	Ratio	Relative freq. (per 10000 words)
<i>mees</i> -ending	626	75	95%	7.47	~20 000	~100	98%	1
<i>naine</i> -ending	35	12	5%	0.42	~400	~20	2%	0.02
<i>mees</i> -beginning	79	17	40%	0.94	~1500	~30	35%	0.07
<i>nais</i> -beginning	117	26	60%	1.40	~2900	~50	65%	0.14

The frequencies of the Estonian Reference Corpus have been rounded to the nearest hundred, as the larger size of this corpus along with the manual revision and calculation of frequencies makes presenting the exact data challenging due to potential spelling variations. The data reveals that across all examined corpora, there is a significant quantitative asymmetry in both token and type frequencies of compound words ending with a gender marking, with more words referring to men. For instance, the token frequencies of compounds ending with *mees* account for 95% of compounds ending with *mees* and *naine* in the sports news corpora, as well as 98% in the Estonian Reference Corpus. In contrast, the token frequencies of words beginning with a gender marking show the opposite pattern, with words referring to women appearing at higher frequencies.

Given the different sizes of these corpora, I calculated relative (i.e., normalized) frequencies using an equation by Brezina (2018: 43), with frequency per

10 000 words as the normalization base. Specifically, I divided the token frequency of each word group by the total number of words in the corpus and then multiplied it by 10 000. As for the sports news corpora, the total word count represents the combined sum of words from both Delfi and ERR corpus. The results showed that relative frequencies were higher in the sports news corpus. While the different sizes as well as represented registers of the examined corpora have an influence on the outcomes, this nevertheless indicates that gendered vocabulary is overall more prevalent in sports media. At the same time, words with a feminine gender marking are used disproportionately less in both corpora, illustrating a male-centric language use regarding compounds with a gendered base form.

While the table summarizes frequencies in both Delfi and ERR corpora, I also analyzed them separately. The frequencies were fairly similar, with compounds ending with *mees* dominating in Delfi (394 occurrences in Delfi and 232 in ERR) and those ending with *naine* in ERR (14 occurrences in Delfi vs 21 in ERR). Hence, regarding words ending with a gender marking, it appears that the texts of ERR are a bit more inclusive with regard to women. As for words beginning with a gender marking, these have also been used more frequently in ERR, with 68 occurrences of words beginning with *nais* and 49 of words beginning with *mees*, compared to 49 and 30 occurrences, respectively, in Delfi.

According to the results of a Chi-square test, the differences in frequencies between *mees*- and *naine*-ending compound words in the sports news corpora and the ERC are also statistically significant (Pearson's  $\chi^2 = 30.1$ ,  $df = 1$ ,  $p\text{-value} < 0,00001$ ). This indicates a noteworthy distinction in how these words are used, indicating a connection between the frequencies of words and the corpus in which they appear. Certain factors differ between these two examined corpora, as the ERC is more heterogeneous. These differences may influence the usage of these words, leading to variation in their frequencies. Furthermore, sports news corpora include texts from only two periodicals, while the Estonian Reference Corpus contains a wider selection of texts. In contrast, the differences in frequencies between compound words beginning with *mees*- and *nais*- in the two corpora were not statistically significant (Pearson's  $\chi^2 = 2.59$ ,  $df = 1$ ,  $p\text{-value} = 0.1073$ ). This suggests that the usage of these word groups does not differ meaningfully in the contexts examined.

Table 4 displays the frequencies of gender-marked words from web corpus 2021 that represent a broader context.

**Table 4.** Frequencies of gender-marked vocabulary in web corpus 2021

Compound group	Token freq.	Type freq.	Ratio	Relative freq. (per 10000 words)
<i>mees</i> -ending	~ 431 200	~ 3500	98%	6
<i>naine</i> -ending	~39 300	~ 730	2%	0.54
<i>poiss</i> -ending	~31 100	~1600	72%	0.43
<i>tüdruk</i> -ending	~12 200	~850	28%	0.17
<i>onu</i> -ending	~6100	~700	48%	0.08
<i>tädi</i> -ending	~6500	~830	52%	0.09

The most frequently occurring words here are compound words ending with *mees*, as they have the highest token frequencies, but also represent the highest variety of different words, given that the type frequency is also the highest for this group. This is a plausible result, as the named words are the most lexicalized in Estonian, representing generic masculinity as well as a range of official occupational titles (such as *esimees* ‘chairman’), and perhaps emphasizing gender less than the other examined compound words. In addition, for compounds ending with *poiss* and *tüdruk*, which refer to younger females and males, the words denoting men (i.e., *poiss*-compounds) have a higher frequency. Words ending with *tädi* occur slightly more frequently than *onu*-compounds in the corpus, and also represent a larger number of distinct words. Compared to the frequencies of other examined word groups, this was the only group where the ratio was almost equal, if not even moderately skewed toward words referring to women. Relative frequencies, which also show frequency per 10 000 words here, display the significant prevalence of *mees*-ending compounds.

## 5.2. Changes in the vocabulary – results of feminist language planning?

Comparing the results of frequency analyses from the corpora of different time periods suggests that the use of marked vocabulary has changed moderately over time. For instance, the ratio of compound words ending with *naine* in the more recent sports corpora is slightly higher than in the ERC representing the 1990s and the 2000s, while the ratio of compounds ending with *mees* has decreased. Additionally, there is a noticeable increase in the proportion of compound words beginning with *mees*. However, the balance of relative frequencies presents a different picture. Relative frequencies of all sports-related gender-marked words are higher in the more recent corpus (see Table 3 above). For instance, while the ratio of compounds beginning with *mees* in token frequency terms has risen 5% in the newer corpus – suggesting that marking male individuals with gendered

initial forms might be increasingly common – the occurrence of compounds beginning with *nais* remains consistently higher, with twice as many compounds beginning with *nais* per 10 000 words than those beginning with *mees* across both corpora. This indicates no significant change over time in the lexical balance referring to women versus men, and suggests that women remain the “unexpected” category. The purpose of gendered initial form in a compound is to denote individuals who work in professions traditionally considered incongruent with the stereotype of their gender, signaling a deviation from the presumed norm (Romaine 2001: 158; Puna 2006: 75; McConnell-Ginet 2020). Thus, the predominance of compound words beginning with *nais* in sports news seems to confirm that an athlete, for example, is typically assumed to be male, while a female athlete is viewed as a deviation from the norm, necessitating more frequent labeling. When adding the data for compounds ending with *mees* and *naine* from the 2021 web corpus, the token frequency ratios are identical to those in the ERC. However, frequencies for both compounds ending with *naine* and *mees* per 10 000 words are higher in the web corpus than in the ERC. Relative frequency of compounds ending with *mees* is lower in the web corpus than in the sports news corpora, while the relative frequency of compounds ending with *naine* is higher in the web corpus.

Hence, the sports news corpora have the most generic masculines in their vocabulary, the use of which seems to have increased over the years. There is also a slight shift towards marking male agents with a gendered initial form. This suggests that male roles might be increasingly specified in the sports context, but the overall balance still emphasizes females more often as the marked category. Furthermore, as the web corpus includes the most unedited new media texts and also the highest frequency per 10 000 words of compounds ending with *naine*, one might speculate whether this reflects a shift toward increased visibility of women in the vocabulary. This could simply indicate that compounds ending with *naine* are more common in colloquial language, which is prevalent in the new media texts. On the other hand, the slightly higher frequencies of compounds ending with *naine* in the sports news corpus than in the ERC may also reflect an increase in the number of female agents (in this context, athletes) being discussed in the texts.

It is difficult to draw definite conclusions about the influence of feminist language planning based on these relatively minor changes in the frequencies. In terms of the frequencies of gendered vocabulary in sports context, language planning does not appear to have significantly impacted Estonian language and, more specifically, sports media language, as men continue to be the default and the expected gender in vocabulary over the years. When I also examined gender-neutral counterparts for several masculine terms in the sports news, I found that some of the neutral words were used more often – for instance, *mängija* ‘player’ (87%) versus *mängumees* ‘playerman’ (13%), *rallisõitja* ‘rally driver’ (58%) versus *rallimees* ‘rally man’ (42%), *tulevikuloetus* ‘future hope’ (76%) versus *tulevikumees* ‘future man’ (24%). However, these are examples of more formal vocabulary which may not be perceived as alternatives to masculine forms (Le Lamer

2012) but rather as standalone terms, and it is unlikely that journalists have chosen them specifically to be gender neutral. There has been almost no public discussion on feminist language planning in the Estonian language and therefore it is unlikely to have influenced the linguistic choices of journalists.

While the frequencies of gender-marked vocabulary illustrate minor changes in biased representations of men and women in language, such biases have perhaps receded more notably in usage. An example of this is the word *esinaine* ‘chairwoman’. Until 2022, language planners recommended to use *esinaine* solely for referring to the chair of an all-female organization (Mäearu 2008). However, an analysis of its usage throughout the century presented in P2 revealed different patterns in more modern language, suggesting that this recommendation may have been relevant only a hundred years ago. For instance, in materials from 1920–1949, *esinaine* in almost all analyzed instances (95% of sentences across three decades) referred to a leader of an all-female organization. Between 1950–1979, its usage was evenly split between the two meanings – chair of an all-female vs. chair of mixed-gender organization. In the more recent texts, from 1990–2008, 2013 and 2021, *esinaine* referred to the leader of an all-female organization in 32% of the examples, while in 66% of the examples, the leader of a mixed-gender organization was denoted (in 2% of the examples, the type of the organization mentioned in the sentence was not detectable). These findings are reasonable, as during 1920–1979, women had little access to leadership positions in mixed-gender organizations. Several women’s unions, active primarily in the early 20th century, united women in Estonia to enhance their influence and to promote politics and culture (Karro 2022), as this was the only way for women to engage in such social activism. *Esinaine* in the materials of 1920–1949 referred to the leaders of these unions. In contrast, today, women lead all kinds of collectives, making this recommendation problematic in many ways. This issue is further exemplified by other languages such as French that have had recommendations to use masculine terms even when women make up the majority of a group (Tudisco 2021). Not only did the studied recommendation not reflect actual language use, but it also reinforced outdated sexist notions that women can only lead other women, and that leaders of mixed-gender organizations are self-evidently male. Since the word *esimees* ‘chairman’ was recommended for use instead, it implies that women needed to be referred to by a masculine term when working in a field perceived as traditionally masculine, such as leadership.

In March 2022, in the Estonian Language Institute’s dictionary *Sõnaveeb*, this recommendation was updated, and the usage of *esinaine* is now described as ‘an elected female manager of a company, institution or organization’. Although this change was not directly influenced by feminist language planning, given the lack of related discussions, it still aligns with the principles of such activism in the domain of dictionary work. As dictionaries hold considerable influence by shaping how words are understood and used by speakers (Cameron 2024: 77), updating such recommendations is a step towards eliminating damaging effects of sexist language. This example does not imply that language itself is inherently sexist; rather, it reflects a sexist context in which these linguistic elements are assigned

meaning. Secondly, naming women explicitly in vocabulary and using such vocabulary productively is necessary, according to the principles of feminist language planning, as it promotes feminization as one strategy for eliminating bias. *Esinaine* exemplifies this approach, as it derives from the male-indicating term *esimees*. The recommended usage highlights a certain stigma attached to vocabulary referring to women in Estonian, as also evidenced by the low frequency and rarity of *naine*-ending compounds. Such compounds have historically not been lexicalized, since naming men in various roles has been more common. Today, however, women are increasingly present in a wider range of roles, including traditionally male-associated ones. A certain pattern has therefore emerged: previously, there were not enough women in certain roles to warrant naming them explicitly. Now, as women increasingly enter these fields, the vocabulary needed to describe them is insufficient. Therefore, the changing usage of *esinaine* represents breaking a certain barrier in the traditions of naming women in language.

Thus, there are signs of feminization of the vocabulary in the Estonian language. This is exemplified by *esinaine* along with similar words like *ärinaine* ‘businesswoman’, *jahinaine* ‘huntswoman’, and *turvanaine* ‘security woman’<sup>13</sup>. Although words referring to women are still used with very low frequencies compared to words referring to men, their usage has been growing slowly – for instance, the absolute frequency of the word *jahinaine* ‘huntswoman’ is 77 times higher in the Estonian National Corpus 2023 (1005 occurrences) than in the ERC from 1990–2008 (13 occurrences). This may reflect a rising number of female hunters, however, linguistic naming and increasing participation of women are mutually influencing factors. Given this trend, feminization could be a relevant language planning strategy for Estonian, as it has already emerged in certain terms, could gradually make women more visible and reduce the stigma around *naine*-ending words. However, attitudes toward feminized words vary among women, based on my personal observations. For example, some women report feeling uncomfortable with being called *esinaine*, seeing it as overly gender-specific, while others prefer it over *esimees*. This illustrates a drawback of feminization, especially as contemporary understandings of gender move beyond a binary framework. Gender-specific vocabulary is not inclusive of diverse identities. In Estonian, there have been few deliberate efforts to create gender-neutral terms, aside from the 2022 recommendation in the combined dictionary of the Estonian Language Institute (ÜS) to use gender-neutral synonyms for *esimees* (such as *spiiker* or *eesistuja*) and the European Union’s gender-neutral language guidelines. Notably, most gender-specific words in Estonian pertain to occupations, such as *esinaine* and *esimees*, where reform efforts have focused the most. It is difficult to assess how language users have been influenced by feminist language planning and whether these guidelines have reached them, as no research has been conducted on its impact or on attitudes toward it. Due to the

---

<sup>13</sup> These words appear among the 55 most frequent compounds ending with *naine* in the Estonian National Corpus 2023

insufficient discussions on this topic, there is also little awareness of feminist language planning. Drawing on personal experience, many language users appear to be open to adopting more neutral and/or inclusive language, and understand the need to implement such guidelines. This perspective emerged from some of the survey feedback. However, opposing viewpoints exist as well, with many adhering to the view that Estonian is already gender neutral. This can also be exemplified by survey comments, such as one participant stating that Estonian should not be “polluted with political correctness”. Conducting a follow-up study on the reception of feminist language planning would therefore be a valuable avenue for further research.

### 5.3. Semantic categories of gender-marked words

Gender-marked vocabulary in Estonian conveys various categories that are introduced and described in this subsection, to address research question 3. The semantic categories are primarily examined on the example of compounds ending with *tädi* ‘aunt’ and *onu* ‘uncle’ in P3, as well as other word groups, such as compounds ending with *mees* and *naine* in P1. The category of occupations in this subchapter is analyzed in greater detail, to also answer research question 4.

**Occupations and activities.** In Estonian, gendered vocabulary most often pertains to different occupations and activities. Specifically, words ending with a gender marking often denote (official) occupational titles – such as *jahimees* ‘hunter’, *põllumees* ‘a farmer’, *kaupmees* ‘a merchant’, *pillimees* ‘an instrument player’, *esinaine* ‘chairwoman’, *ärinaine* ‘businesswoman’, *jahinaine* ‘hunter-woman’, *ilmatüdruk* ‘weather girl’, *pallipoiss* ‘ball boy’. These terms are the most lexicalized, as mentioned above. Additionally, many words ending with a gender marking also describe men or women engaged in specific hobbies or activities – such as *tsiklimees* ‘motorcycle man’, *suitsumees* ‘smoking man’, *käsitöönaine* ‘handicraft woman’, *trennina* ‘workout woman’, *tantsutüdruk* ‘dancing girl’, *kooripoiss* ‘choir boy’, *joogatädi* ‘yoga aunt’, *korvionu* ‘wicker basket maker, lit. wicker basket uncle’. For words that begin with a gender marker, occupations and role nouns also form the largest category. Here, gender markers differentiate the gender of the representative of a certain profession, as in *meesõpetaja* ‘male teacher’, *meesartist* ‘male artist’, *meesarst* ‘male doctor’, *naisnäitleja* ‘female actor’, *naistöötaja* ‘female worker’, *naisjuht* ‘female leader’.

Various subcategories of words referring to occupations and activities can be distinguished. For example, sports-related compound words refer to an athlete primarily through specific sports (e.g., *hokimees* ‘hockey man’, *rallimees* ‘rally man’, *laskesuusana* ‘biathlon woman’, *epeenaine* ‘épée woman’). This category was the most prevalent among compound words ending with *mees* in sports news, accounting for 41% of all compounds ending with *mees*. In contrast, for compound words ending with *naine*, this category contained the highest number of distinct words, though they were less frequent overall, making up 17% of all words ending with *naine*. Additionally, these compound words also defined male

and female athletes through achievements (e.g., *rekordimees* ‘record man’, *võidumees* ‘winner man’, *pronksinaine* ‘bronze (medal) woman’), positions (*keskväljamees* ‘midfielder man’, *ründamees* ‘striker man’, *ankrunaine* ‘anchor woman’), other activities (*määrdeemees* ‘ski waxer, lit. waxer man’, *vilemees* ‘referee, lit. whistle man’), and future potential (*tulevikumees* ‘future man’). Compound words beginning with *mees* and *nais* mostly referred to specific sports (*meessportlane* ‘male athlete’, *meesujuja* ‘male swimmer’, *naissportlane* ‘female athlete’, *naisjooksja* ‘female runner’), in addition to age group, group membership or competence (*naisproff* ‘female professional’). Many compounds beginning with a gender marking were used in contexts like awarding the best athletes of the year, where gender differentiation in language may hold some significance. Interestingly, there is a difference of opinion within the Estonian feminist community<sup>14</sup> about whether women and men should be recognized separately for awards – especially regarding whether, for instance, outstanding women in traditionally male fields should be awarded independently.

Upon closer inspection, it is evident from this vocabulary which sports are more frequently associated with men and which with women. For example, the lion’s share of compound words ending with *mees* referred to ball games, such as football, as illustrated by the most frequent word *mängumees* (‘player, lit. game man’). Additionally, there were more words related to rally and motorsports. In contrast, while there were fewer compounds ending with *naine* in this category, skiing appeared to recur more frequently among them. Therefore, since sports are traditionally viewed as a masculine field, it can be concluded that the aforementioned sports are particularly associated with men. However, these results may also reflect the fields in which Estonian athletes were the most successful at the time. The frequencies of compounds ending with *naine* were, however, too low to draw similar conclusions.

Indeed, the gender-marked vocabulary allows conclusions to be drawn about which occupations are associated with men and which with women. This is further highlighted by occupational terms ending with *tädi* and *onu*. While these words do not represent the official occupational titles but colloquial language, they nonetheless shed some light on the roles associated with women and men. In general, *tädi* in occupational titles primarily marked professions related to customer service, (e.g., *müüjatädi* ‘seller aunt’, *vastuvõtutädi* ‘reception aunt’, *piletitädi* ‘ticket aunt’), including cleaning services (*koristajatädi* ‘cleaning aunt’, *pesumajatädi* ‘laundromat aunt’), healthcare (*arstitädi* ‘doctor aunt’, *haiglatädi* ‘hospital aunt’), education and social work (*kasvatajatädi* ‘kindergarten teacher aunt’, *sotsiaaltädi* ‘social worker, lit. social aunt’). In contrast, *onu* predominantly represented law enforcement (*politseionu* ‘police uncle’, *turvaonu* ‘security uncle’), healthcare (*arstionu* ‘doctor uncle’) and customer service (*müüjaonu* ‘seller uncle’, *kassaonu* ‘cashier uncle’). Additionally, *onu* appeared in words

---

<sup>14</sup> There was a discussion in the Facebook group of Estonian feminists about whether a national science competition of high school students should include the “Best Girl” award among the overall prizes.

marking occupations such as construction, business, entrepreneurship, and science (*elektrionu* ‘electrician uncle’, *corp-onu* ‘corporate uncle’, *teadlaseonu* ‘scientist uncle’) – fields which did not emerge with *tädi*-compounds. Such biases have been found to be propagated by large language models, which were found to exhibit stronger bias scores for words referring to either men or women (e.g., *tädi* or *onu*), aligning with human biases about certain occupations as reflected in labour force statistics (Kaukonen et al. 2024b). Furthermore, compounds indicating male individuals frequently referred to criminality, reinforcing the stereotypical assumption that men are more prone to committing crime. For instance, the most frequent *onu*-compounds was *kommionu*, ‘lit. candy uncle’, a jocular and somewhat softening reference to pedophile. Similarly, words such as *pommimees* ‘bomb maker, lit. a bomb man’ and *vargapoiss* ‘thief boy’ were also more frequently observed than their feminine counterparts. According to the vocabulary, men were also more often linked to activities such as drug dealing or use, as comparable terms referring to women did not emerge. Drinking alcohol, however, appeared to be more frequently associated with women. It is possible that the word *joodik* ‘drunkard’ generally refers to men, making *joodiknaine* ‘drunkard woman’ more notable. In contrast, words referring to women were more likely to describe involvement in sex work (*lõbunaine*, *lõbutüdruk* ‘a sex worker’), a theme that was less common in compounds referring to men. Examples of the latter include *pornomees* ‘male porn actor, lit. porn man’.

Thus, Estonian gender-marked vocabulary reflects stereotypes and traditional gender roles, highlighting real-life gender segregation in employment: women are predominantly found in fields such as healthcare, social welfare, education, and service sector, while men are overrepresented in construction (Statistikaamet 2022). Moreover, language also conveys information about the relative prestige associated with specific occupations and, by extension, their representatives. *Tädi* often represents professions of lower prestige, affirming that female-dominated professions, such as education and care work, are undervalued as well as underpaid (England et al. 2002; Parent-Lamache & Hallée 2023). This could be further illustrated by, for instance, the seemingly equivalent words *pangaonu* ‘bank uncle’ and *pangatädi* ‘bank aunt’ that differ in meaning: *pangaonu* primarily referred to a banker, while *pangatädi* to a teller. This aligns with Kerli Puna’s (2006) findings, which suggest that female representatives of high-status professions are rarely referred to as “aunts”, as confirmed by words such as *äritädi* ‘business aunt’ or *teadlasetädi* ‘scientist aunt’ that occurred very infrequently in the materials of this study.

The notion that terms referring to women (e.g., *tädi*-compounds) denote lower status is further supported by wage disparities – for instance, kindergarten teachers earn about 440 euros less than district police officers, according to the second-quarter 2024 statistics (Statistikaamet 2024b, Rahandusministeerium 2024). Furthermore, *tädi*- and *onu*-compounds reflect the semantic pejoration described by Schulz (1990 [1975]), with *tädi*-compounds often carrying more negative connotations. While *tädi*- and *onu*-compounds are frequently used to denote individuals working with children – with the aim to express familiarity or warmth –

their usage changes in adult-to-adult contexts, which do not involve children. In this case, terms drawn from children's language, like *tädi*, are often used jocularly or ironically: the study of their usage revealed that the usage examples of *tädi*-compounds were frequently employed to ridicule or mock the female individuals referred to as *tädi*. What is more, both *tädi*- and *onu*-compounds can also serve derogatory purposes, mocking or belittling the individuals they describe and conveying dissatisfaction with or criticism toward them (P3: 156–157, 159–161). *Tädi*-compounds, in particular, are frequently used to belittle women by assigning them a lower status. Jurafsky (1996) discusses analogous effects regarding diminutives, which, in addition to their affectionate connotations, are often used to mark socially marginalized female gender. The semantics of diminutives is frequently associated with both children and women, and in such cases, they often carry pejorative meanings. The root of this, according to Jurafsky, lies in diminishing women and equating them with children. *Tädi* can be a similar example of this pejoration, as it also often links women with childcare and other communal roles.

Another noteworthy finding about Estonian gendered occupational titles is that gender markings are productively used in word formation, enabling almost any occupation to be expressed with a gendered term, and contributing to the formation of new words. This is particularly evident for compounds ending with *mees*, where the gender emphasis has diminished the most due to their generic usage. For example, alongside the more frequent and established terms ending with *mees* in sports news, several more striking examples emerged, such as *sangpommimees* 'kettlebell man', *halloomees* 'anchorman, lit. hello man', *võhmadees* 'lit. stamina man'. It is likely that the originators of these words intended to refer specifically to men, possibly aiming to use more colloquial language in media texts, as gender-neutral terms might have been more natural in these contexts. Thus, while such new formations are relatively informal and may emphasize gender more strongly, their creation ultimately reinforces the genericity of men. Another example of gendered terms used to create occupational titles are *tädi*- and *onu*-compounds. These colloquial terms can refer to representatives of various professions, while also being strongly gendered and conveying additional meanings – such as age, appearance, or behavior. *Tädi* and *onu* often denoted specific occupations, such as *maasikatädi* 'strawberry seller,' lit. 'strawberry aunt', *kindlustusetädi* 'insurance office worker, lit. insurance aunt', *parkimis-onu* 'parking lot security guard, lit. parking uncle', *trepionu* 'staircase repairer, lit. staircase uncle'. While writing this thesis, I could refer to myself as *väitekirjapädi*, i.e. 'dissertation aunt'. In this context, words referring to occupations and activities often describe male and female individuals through metonymy.

**Age.** Estonian gender-marked vocabulary frequently identifies male and female individuals through age (e.g., *vanamees* 'old man', *vanaperenaine* 'old landlady', *vanapädi* 'old woman, lit. old aunt', *täismees* 'grown man', *beebitüdruk* 'baby girl', *koolipoiss* 'schoolboy'). While compounds indicating age do not have the highest type frequencies – suggesting there are relatively few different words with gender markings to denote age – the ones that do exist are among the most

frequently used. Additionally, gender marking in the initial form of a compound word can modify the main word denoting age, as seen in terms used in sports news (*naisjuunior* ‘female junior’, *meesseenior* ‘male senior’), which indicate age and experience groups in sports. Several gendered words inherently denote age, such as *tüdruk* and *poiss*, *tädi* and *onu*. The latter two, *tädi* and *onu*, are generally used to address older individuals in late adulthood or the third age (i.e., starting from 60 years). As a result, these terms can have a derogatory effect when used to refer younger people to due to age-related implications. Specifically, *tädi* is often employed in jokes based on stereotypical female characteristics and aging, serving as both a critique of aging women and a warning to younger women about aging and certain lifestyle or appearance choices associated with it (Lövgren 2013: 133; Naistekas 2006). This reflects societal views that tend to perceive women’s aging more negatively than men’s. However, based on personal observations, young women on social media are increasingly embracing and glorifying a calmer, so-called “grandmotherly” or “aunt-like” lifestyle, challenging its traditionally negative associations. This trend, however, would require more systematic research – likely in anthropology or cultural studies rather than in linguistics – to be thoroughly analyzed and validated.

**Relationships.** Another semantic category that emerged encompasses both close and distant relationships, including family relations (*vanatädi* ‘great aunt’, *vanaonu* ‘great uncle’, *vennanaine* ‘sister-in-law’, *naabrimees* ‘neighbour man’, *peigmees* ‘groom, boyfriend’, *väimees* ‘son-in-law’, *kolleegitädi* ‘colleague aunt’). Within this category, several words also denoted marital status, such as *eksmees* ‘ex husband’, *abielunaine* ‘married woman’, *poissmees* ‘bachelor’, *vanatüdruk* ‘spinster’.

**Cultural and social identity.** The semantic category of cultural and social identity encompasses compound words that first function as geographic classifiers. Specifically, the initial forms of these gender-marked compound words referred to nationality (*lätipoiss* ‘Latvian boy’, *venemees* ‘Russian man’, *ameerikamees* ‘American man’, *hiinaonu* ‘Chinese uncle’, *saksatüdruk* ‘German girl’, *poolatädi* ‘Polish aunt’, *latiinonaine* ‘Latina woman’, *koreapoiss* ‘Korean boy’). A significant subset of these words also includes those denoting ethnic groups, such as *berberinaine* ‘Berber woman’, *eskimonaine* ‘Eskimo woman’, *abori-geeninaine* ‘Aboriginal woman’, *mustlastädi* ‘gypsy aunt’, *mustlaspoiss* ‘gypsy boy’, *indiaanitädi* ‘Native American aunt’, *indiaanimees* ‘Native American man’, *beduunimees* ‘Bedouin man’. Additionally, words referring to race are also included in this category. The category further includes compounds expressing place of residence. Examples include *külamees* ‘village man’, *linnatädi* ‘city aunt’, *pealinnatüdruk* ‘capital girl’, *hiiumees* ‘Hiiumaa man’, *tartupoiss* ‘Tartu boy’, *välismaamees* ‘foreign man’. Second, words expressing religion also fall into this category, including *juudinaine* ‘Jewish woman’, *islamionu* ‘Islamic uncle’, *moslemine* ‘Muslim woman’, *moslemitädi* ‘Muslim aunt’, *kirikumees* ‘church man’. Similarly, words referring to gender and sexuality are represented (*geimees* ‘gay man’, *lesbitädi* ‘lesbian aunt’, *vikerkaaremees* ‘rainbow man’,

*homopoiss* ‘homo boy’, *transnaine* ‘trans woman’, *binaine* ‘bisexual woman’, *cis-naine* ‘cisgender woman’, *meesnaine* ‘transgender woman, lit. manwoman’).

Importantly, many of these terms carry connotations that are neither neutral, objective, nor accurate. Instead, they are often outdated or derogatory labels. This highlights how language not only encodes gender bias but also reflects discriminatory and prejudicial attitudes toward individuals of different sociocultural identities, highlighting historical and social biases and stereotypes.

**Physical and behavioral traits.** Gender-marked compounds describe male and female individuals through appearance, including clothing, hairstyles, and physical shape. This is illustrated by words such as *pakspoiss* ‘fat boy’, *kaunis-meess* ‘pretty man’, *vuntsionu* ‘moustache uncle’, *kikilipsuonu* ‘bow tie uncle’, *käabusnaine* ‘midget woman’, *hiiglasnaine* ‘giant woman’. This category also includes terms related to specific subcultures (*emotüdruk* ‘emo girl’, *hipionu* ‘hippie uncle’, *rastaonu* ‘rasta uncle’, *rokinaine* ‘rock woman’, *ossmees* ‘chav man’). Some unique and creative examples stand out in this category, for instance, *külmikonu* ‘refrigerator uncle’, which describes a muscular man resembling a fridge; *pandatädi* ‘panda aunt’, referring to a woman with dark circles around her eyes; *plastiknaine* ‘plastic woman’, a woman who has undergone extensive cosmetic surgeries; or *notsutädi*, ‘pig aunt’, a woman who has pig-like appearance. It is also notable that words denoting women are often used negatively, frequently serving as tools for criticism.

Behavioral traits and characteristics are also expressed through the vocabulary (*vampnaine* ‘vamp woman, femme fatale’, *machomees* ‘macho man’, *pehmomees* ‘softie man’, *power-naine* ‘power woman’, *psühhonaine* ‘psycho woman’, *päikesetüdruk* ‘cheerful girl, lit. sun girl’). Additionally, compounds often reflect associations with zodiac signs as well as the act of driving certain cars, especially for describing men (*mersumees* ‘Mercedes Benz man’, *hondamees* ‘Honda man’, *džiibinaine* ‘Jeep woman’, *bemarimees* ‘BMW man’, *avensise-meess* ‘(Toyota) Avensis man’, *audi-poiss* ‘Audi boy’). In sum, it is evident that similarly to occupational titles, almost any physical or behavioral attribute can be represented through a gendered compound word. These compounds also often reflect and reinforce gender stereotypes, and can carry negative connotations.

**Non-human entities.** Interestingly, many gendered compounds refer to non-human entities – more specifically, animals, creatures, characters and inanimate phenomena. For example, the names of various animals include (primarily masculine) gender markings, such as *põdraonu* ‘uncle Moose’, *hiirepoiss* ‘mouse boy’, *tihasetüdruk* ‘titmouse girl’. This tendency is derived from the children’s book “The Tales of Uncle Remus” (from Julius Lester), where the characters in the Estonian translation were named uncles. During the Soviet era, the recording of the work was released in Estonian, which became very popular. Hence, these stories may have had a cultural impact on language use. Compound words referring to animals might also include gender markings, such as *onu*, *poiss* or *tüdruk*, due to their child-like associations, as these convey playfulness, informality and non-threatening qualities. Similarly, the more informal words for many inanimate objects and natural phenomena include masculine gender

markings, such as *vässuonu* ‘lit. uncle tiredness’ (as in ...*ma ei läinud Rabakale, vässuonu on...* ‘I didn’t go to Rabaka, I have uncle tiredness’), *viinaonu* ‘vodka’ (as in *Avasime muuseas just viinaonu...* ‘We just opened a vodka uncle...’), *pisarapoiss* ‘lit. tear boy, a tear’, *tuulepoiss* ‘lit. wind boy, wind’.

**Names and usernames.** Additionally, gender markings are often attached to proper names, such as *Jaanpoiss* ‘boy Jaan’, *Rita-tädi* ‘aunt Rita’, *Villuonu* ‘uncle Villu’, as well as various usernames in the forums that are included in the materials of the corpus.

## 5.4. The impact of language and stereotypes on gender perception

With this study, I aimed to investigate whether the perception of occupational titles is influenced by language or social knowledge and whether generically used occupational titles are truly perceived as such. The results suggest that both language and social knowledge influence this perception, illustrating the role of language in perpetuating stereotypes.

The results of the survey, which was designed to investigate gender perception (research question 5) revealed that the majority of Estonian occupational titles without a morphological gender marking are perceived as neutral. This indicates that the language users do not associate job announcements with a representative of a specific gender. Neutral responses were, in fact, the most common regardless of whether the occupation exhibited gender bias in real life. However, there were exceptions: for a few strongly biased occupations in real life (e.g., *sõdur* ‘soldier’, *küünetehnik* ‘nail technician’), perceptions were also skewed, with these being predominantly associated with either males or females. In contrast, the responses for gender marked occupational titles were biased towards the corresponding marker. Thus, words with a feminine and masculine gender marking result in predominantly female- and male-biased responses, respectively (P4). Additionally, the effect of lexicalization on interpretation was evident for gender-marked words: the more frequently used and thus more entrenched the gender marking in the occupational title, the more neutrally it is perceived. This was demonstrated by words such as *esimees* ‘chairman’ and *medõde* ‘nurse’, which received more neutral responses despite their gender marking. This neutral perception might also be influenced by the blurring of gender boundaries, which is more apparent in certain professions than others. Nevertheless, the distribution of responses further revealed that words denoting generic masculinity exhibit a greater bias compared to their neutral counterparts. Furthermore, some titles with feminine gender marking showed even greater bias than titles with a masculine gender marking, as these are not used generically. Therefore, interpretations seem to be shaped both by social reality and stereotypical thinking, the latter of which is further influenced and reinforced by linguistic cues.

We additionally found that Russian language speakers tend to rely more heavily on stereotypical associations, with stereotypes often overriding linguistic

information, consistently with earlier studies (Kapatinski 2006; Garnham & Yakovlev 2015). For example, several female-stereotyped occupational nouns (e.g., *мастер маникюра* ‘nail technician’, *воспитатель* ‘kindergarten teacher’) dominantly elicited female-biased responses in the Russian survey, despite these words being presented in a grammatically masculine form. In addition, the results revealed that Russian grammatically feminine words are largely perceived as female-biased, while grammatically masculine words typically elicited neutral responses. While grammatically masculine words in Russian specify gender, these results reveal that they are less marked than feminine words and therefore convey less explicit gender information – both linguistically and socially, as their gendered meaning has diminished due to their generic usage. In contrast, feminines emphasize (female) gender, and elevate its significance. However, some exceptions emerged here as well, further demonstrating the interplay between language and stereotypical assumptions in shaping gender perceptions. For instance, for the word *модель* ‘model’, while being grammatically feminine, nearly half of the responses were neutral. This probably reflects its lack of a typical feminine marker and the fact that modeling is a gender-balanced profession in Russian society. Overall, Russian operates differently in gender expression than most other grammatically gender languages, since in Russian, grammatical gender does not include agreement with articles. Instead, gender is expressed by suffixes in nouns, and the gender of a noun determines the form of adjectives and verbs accordingly, which change to match the gender of the noun they are associated with. Garnham and Yakovlev (2015) also highlight the uniqueness of Russian in comparison to other grammatically gendered languages, such as French or German. Speakers of these languages have been found to rely on grammatical cues even in situations where the stereotype of the role noun mismatches its grammatical gender (Gygax et al. 2008). This has not been observed among Russian speakers. This highlights how various forms of linguistic gender markings affect gender perception in distinct ways.

Grammatically feminine words aside, the majority of responses for the examined occupational titles were perceived neutrally in the Russian survey as well. The central tendency bias, which often characterizes similar surveys, needs to be considered here (see Chapter 3.2.). It is possible that the majority of participants in our surveys did not perceive these occupations as predominantly male- or female-dominated. Moreover, since we explicitly instructed participants to adopt the perspective of employers, it is also possible that the judgments might reflect the assumption that employers should act impartially during a recruiting process, based on social expectations. Another possible explanation of such results could be the level of exposure. The more exposure individuals have to the representatives of a certain profession, the more stereotypical information has accumulated. Hence, since some of the professions are encountered infrequently by most individuals, gender-stereotypical information about representatives of these professions may be limited, and therefore such professions might not be perceived in a biased manner. On the other hand, perceptions may also derive from access to stereotypical information only. Despite these factors, the large sample size for

both these surveys enables us to draw some conclusions and make some generalizations (Andrade 2020).

Statistical analyses conducted further demonstrated the influence of both language and stereotypes on gender perception. In the principal component analysis (PCA), responses to strongly stereotyped occupations accounted for the largest variance in the data across both surveys. The second principal component was represented by words with an explicit gender marker in Estonian and varying grammatical gender forms in Russian. The third component was primarily illustrated by occupations that were perceived as neutral in both surveys, suggesting that information or belief about gender balance in certain occupations also possibly shapes their perceptions. Distinct clusters in the PCA biplots (P4: p. 7) for Estonian indicated a stronger inter-variable relationship among words without morphological gender markers denoting neutral and masculine occupations, and those with masculine gender markers. Similarly, stronger relationships were observed among words denoting feminine occupations and those with feminine gender markers. In Russian, grammatically feminine words representing female-dominated occupations showed stronger internal relationships, while occupations represented in the grammatically masculine form exhibited distinct relationships within their group (P4: p. 13). Binomial logistic regression models further revealed that occupational stereotypes and morphological gender marking – both feminine and masculine – were strongly associated with biased responses in Estonian. Words with a feminine gender marking were particularly influential in increasing the odds of (female)-biased responses, highlighting the gender-emphasising aspect of these words compared to words with a masculine gender marking. In Russian, grammatical gender and the type of occupation were the most significant predictors of biased responses. Masculine grammatical forms and male-dominated occupations were associated with an increased likelihood of male-biased responses, while feminine grammatical forms and female-dominated occupations were associated with an increased likelihood of female-biased responses. As a result, different analyses conducted provide evidence of patterns in language perception, highlighting the significant interaction of language and social knowledge in forming the meanings and interpretations of words – regardless of the typology of a language. These aspects also interact with each other and can contribute to gender inequalities, the implications of which I will explain in the next subsection. I also acknowledge that the tendencies identified in this thesis could benefit from further exploration through more implicit and nuanced experimental methods, and be confirmed by additional statistical approaches, as I will discuss in Chapter 5.6.

## **5.5. Gender bias in a genderless Estonian**

In this subsection, I will summarize the results highlighted in previous sub-chapters, discussing how Estonian language, which is believed to be gender-neutral, conveys gender bias.

First, several conclusions can be drawn from the frequency analyses. The large predominance of compound words ending in *mees* in the examined materials demonstrates the generic nature of words referring to male individuals, and suggests that men and masculinity are more often equated with universal humanness, as illustrated by the usage of gender-marked vocabulary. Since many professions have traditionally been held by men, such words include a masculine gender marking that reflects this tradition (Puna 2006). The historical dominance of men in most professions has, in turn, led to these terms becoming more entrenched in language and thus seemingly more natural than compounds ending with *naine*. This is illustrated, for instance, by the sports context, the language of which may be particularly male-centric due to the traditional association of sports with men and masculinity. As compound words ending with a masculine gender marking can be used to refer to a person of any gender, they are probably used predominantly in the sports texts due to the writers' aim to denote an athlete in general. Undoubtedly, the specifics of the (sports) news genre and the variation of genres and periodicals in the examined corpora must also be considered here. The predominance of compound words ending with *mees* may stem from journalists' efforts to use more synonyms when referring to an athlete, along with the possibility that men's sports simply may receive more media coverage. Additionally, as most sports journalists are men, they may tend to employ more frequent use of generic masculinity, since men have been found to interpret generic masculine terms more inclusively, perceiving them as referring to themselves (Hamilton 1991; Redl et al. 2021). Consequently, this mindset gets unintentionally expressed through the authors' linguistic choices. To examine the genericity and bias resulting from the overuse of *mees*-ending compound words, a similar analysis could be conducted by using linguistic data representing other fields, or more broadly general language, as opposed to language of specific genres. Furthermore, experimental studies, which I will discuss in Chapter 5.6, could help identify the extent of genericity in such words.

Nevertheless, although it represents a specific field, sports news is also an important part of public discourse, shaping and influencing reality through its language. Therefore, the frequency ratios of gender-marked words in sports news reveal patterns and regularities of language use that reflect the unequal power relations between men and women in the vocabulary, and reusing terms that express generic masculinity reinforces this bias even further. Generic masculines, as Schneider & Hacker (1973) suggest, are a symbol of male dominance and higher power positions, which are reflected in and thereby reinforced by language.

Therefore, Estonian also demonstrates the existence of the *people = male* bias of language to which feminist linguists have brought attention (Cooper 1989; Hamilton 1991; Pauwels 2003; Mills 2008). This could further be illustrated by higher frequencies of *poiss*-compounds, as they can appear to be more gender-neutral than *tüdruk* and are thus used with that intent. As for *tädi*- and *onu*-compounds, these words exhibit the greatest balance of words indicating females vs. males, with compounds denoting females (i.e., *tädi*-compounds) occurring slightly more frequently. This suggests that the word *tädi* and the compounds it is used in

potentially convey a wider range of usages in communication, are more productively employed in word formation than other vocabulary referring to women, and that *onu* is likely not perceived as being as generic as *mees-* and *poiss-* compounds, as *onu* is more informal and refers to a more specific concept. Additionally, while compounds ending with *tädi* and *onu* are used across different genres, it could be argued that they spread more often in a more colloquial, everyday language, that is better represented by the texts of new media (i.e., forums, blogs and commentaries) than edited media texts.

It is also noteworthy that inanimate phenomena in Estonian are named mainly with a masculine gender marking (e.g., with lemmas *poiss* or *onu*) with an aim to achieve a more expressive, humorous and eloquent tone. While the suffix *tüdruk* occasionally appears next to *poiss* in the names of animals to convey youthfulness and playfulness, the suffix *tädi* is absent in such contexts. Moreover, in Estonian, the offspring of animals are also named with a masculine term *pojad* ‘sons’. In the naming of other non-human entities, feminine gender markings are rare. This tendency may also reflect a male-biased perspective, with vocabulary referring to men perceived as more neutral and generic, while female references are seen as more marked or pronounced (Mills 2008: 47). Gender markings in such names personify phenomena, assigning them human-like qualities, with men often perceived as the default representatives of humans.

As I indicated above, the frequency ratios additionally suggest that compound words ending with a feminine gender marking (e.g., *naine-* compounds) are not as deeply rooted and natural in Estonian. In fact, they may even seem uncomfortable, awkward, unusual, and slightly alter the original meaning of the word marking men and place unnecessary emphasis on the gender of women. This phenomenon, along with the rarity of feminine forms and their potential perception as more gender-emphasizing due to language users’ limited exposure to them, has been observed in many other languages as well (Mills 2008; Gabriel et al. 2018; Oksanen 2019; Szuba et al. 2022: 833). Moreover, professions referred to with a feminine marking tend to be assigned a lower status (Merkel 2013). The analysis of the word *esinaine* additionally revealed that while feminine base forms in compound words have gained wider usage over time and some feminine terms have shed their sexist associations, compounds ending with *naine* continue to have a degree of abnormality. In Estonian, however, female agents seem to be more frequently marked with a gendered initial form *nais*, as this is, based on the example of sports context, used more often than gendered base form *naine* and thus appears to be more neutral (i.e., *naissportlane* ‘female athlete’ might be perceived as more formal than *spordinaine* ‘sportswoman’). However, the use of gendered initial forms might also unintentionally make women’s gender more prominent, whilst giving the impression that women are secondary in some specific fields – when talking about *players* and *female players*, for instance.

To provide another example, in the Estonian National Corpus 2023, the word *naisjuht* ‘female leader’ appears 4 times more frequently (about 2700 occurrences) than *meesjuht* ‘male leader’ (about 630 occurrences). Kerli Puna (2006: 65–66, 73) also addressed this phenomenon, noting that women in higher

positions are more often linguistically marked, precisely because they are still considered non-standard and exceptional in such roles (see Marling et al. 2024). As a result, the counterparts of such occupational titles beginning with *mees* (e.g., *meespresident* ‘male president’) would sound strange, according to Puna’s conclusions in 2006. Although Estonia has had both a female president and prime minister, this observation still holds today, and the linguistic domain continues to reflect gender inequality in many areas. Conversely, in certain fields, such as sports, there is also a gradually growing trend of marking men with gender-specific initial forms. This trend can be considered justified, as many roles and occupations have become more gender-balanced than in previous decades. It would be useful to further analyze the usage and perception of such gender-neutral terms (i.e., president, athlete etc.) to examine the distribution of references to women and men. Such analysis would help determine the extent to which gender marking is necessary and to identify which forms of gender marking are appropriate in which contexts, in order to attain language use that is both accurate and sensitive, ensuring that gender information is conveyed only when relevant and in a respectful manner.

The other semantic categories of Estonian gender-marked words also convey bias and stereotypical associations. In addition to traditional gender divisions reflected in the examined words referring to occupations, it is evident that the lemmas marking male gender are more often modified by words referring to criminality, with several examples being lexicalized (e.g., *kommionu*, i.e., *kommi* ‘candy’ SG.GEN + *onu* ‘uncle’ SG.NOM). While, on the one hand, this illustrates the fact that the gender distribution of perpetrators of crime, including murder and sexual offenses, is statistically skewed towards men (Justiitsministeerium 2023), on the other hand, it conveys a prejudice that crime (and violence) is primarily associated with men and masculinity (Naffine 1987: 43). Furthermore, the analysis reveals that compound words with gender markers in Estonian serve not only to specify gender but also to encode other social and cultural dimensions, such as nationality, ethnicity, place of residence, sexuality, religion and many more. However, the choice of terms, particularly those related to ethnic groups, reflects historical biases in addition to gender bias, as well as societal attitudes, with some compounds perpetuating stereotypes and derogatory views. Additionally, women are more often marked in words denoting race, nationality and religion, as the default forms of such words (e.g., *juudinaine* ‘Jewish woman’ vs *juut* ‘a Jew’) refers to male individuals. This further confirms that women represent the non-standard category (Cooper 1989; Pauwels 2003).

In regards to words describing behavioral and physical traits, compound words referring to men and women through various attributes tend to be highly stereotypical, describing men and women in terms of distinctly masculine or feminine traits. Men are more frequently associated with cars and other vehicles, technology, drinking alcohol, being macho, making jokes, liking women, while women are more often linked to being emotional, well-groomed, seductive, being linked to mythical creatures, such as witches, vampires or fairies. Similarly, a study of media representations of male and female leaders (Kaukonen et al. 2024a;

Marling et al. 2024) found that media portrayals reinforce strong gender stereotypes, with men being described through hyper-masculine and women through hyper-feminine characteristics. While the aforementioned study focused on stereotypes reflected in more covert and subtle language patterns, putting emphasis on gendered discourses, these findings align with the strong stereotyping evident also in overt language instances.

Thus, such examples serve as evidence that even a genderless language such as Estonian does not lack gender bias and stereotypes. These can be embedded in the contextual layers of language, but are also reinforced by the language itself. The analyzed web corpus consists of unedited texts, providing an unfiltered overview of how the everyday language perceives women and men, along with the social roles and characteristics attributed to them. It is language users who drive the creation and usage frequency of words, contributing to how certain words become established and lexicalized. Therefore, it is unclear whether this vocabulary reflects inherent qualities of the language itself or the attitudes and beliefs of its speakers. Language and its users exist in a reciprocal relationship and thus language itself *is* a reflection of its speakers and their various cognitive processes, which generate and influence linguistic structures – and vice versa (Croft & Cruse 2003; Boroditsky 2011; Ibbotson 2013). In this way, language acts as a mirror of its speakers who, in turn, mirror the society they inhabit. The observed gender biases and stereotypes in vocabulary therefore stem from social order and social dynamics between men and women, but they are also formed by the cognitive processes of language users that are, in turn, influenced by the surrounding social order. Language is a creation and perpetuation of reality, constantly defined and reproduced by its speakers (Schneider & Hacker 1973: 13), and so is gender, which is being reproduced through language.

Language can, however, also break stereotypes. In the examined vocabulary, several compounds ending with *naine* describe fields and traits traditionally not considered feminine, such as leadership or sports. In addition, various words represent the feminine counterparts of generic masculines, illustrating attempts of feminization in Estonian, as also discussed in section 5.2. Although the frequencies of such words are relatively small, several examples show an increase in their occurrences over the years. The opposite option – breaking the masculine stereotype – appears to be less common, as the masculine stereotype changes are slower.

Gender bias is strongly evident in Estonian gender-marked vocabulary marking occupations and activities, as revealed in the examined corpora. This vocabulary reflects traditional gender divisions and stereotypical associations, as represented in Subchapter 5.3. While the lexicon mirrors actual gender distributions in professions, it also perpetuates societal assumptions and expectations about which professions are considered appropriate – as well as inappropriate – for women and men. Based on this vocabulary, women are typically associated with communal roles that include nurturing and are rarely represented in construction or law enforcement, for instance. Conversely, men are linked to agentic roles that emphasize authority, protection, and technical expertise, while being uncommon

in the educational sector. The use of such vocabulary perpetuates these expectations, which in turn influence real-life scenarios, creating a cyclical relationship between language and societal norms. Moreover, it was evident from the vocabulary that words referring to men and women mark stereotypical activities. For example, women are associated with housework and handicrafts, while men are more often engaged with physical activities.

These findings from the corpus study were supported by the survey results. While most of the occupational titles in the survey were perceived as gender-neutral, biased perceptions persisted, particularly for gender-marked occupational titles in the Estonian survey. The representation and organization of word meanings arise from connections established between words and elements perceived in experience (Bolognesi 2020: 4), which aligns with the social role theory. Hence, the biases stem from real life observations, underscoring the influence of social knowledge on the perception of language units (see Eagly et al. 2000; Eagly & Wood 2012; Koenig & Eagly 2014; Gygax et al. 2016). Moreover, occupational titles including a gender marking generally refer to traditionally masculine or feminine occupations, reflecting the unbalanced gender divisions observed in real life, and resulting in such words being perceived as referring more often to men or women. Such knowledge often serves as the basis of stereotypes. In the case of these survey results, respondents may have not been aware of the actual statistics behind gender divisions in the professions examined in the survey, and instead relied on learned knowledge and stereotypes. These stereotypes are then reinforced by language, which perpetuates existing biases in thought. For example, occupational titles without a morphological gender suffix in the P4 were perceived as more neutral than their gender-marked counterparts containing a masculine suffix. This reveals that while such generic terms have previously been perceived as gender-neutral, they are not interpreted accordingly by language users, and may therefore exhibit bias. Therefore, while biases stem from social knowledge, they are also reproduced by language, aligning with the findings of previous studies concerning other languages (Gabriel & Gygax 2008; Gabriel et al. 2017). These results further indicate that while stereotypes and biases might already exist aside from linguistic cues, the use of neutral terms might still increase the inclusivity and introduce a degree of neutrality.

Occupational gender stereotypes that are reproduced by language are harmful in various ways. First, they reinforce prejudice by consolidating the traditional gendered division of labour and prescribing specific roles for women and men. Such prescriptions can result in preventing individuals from pursuing certain positions – for instance, research has demonstrated that stereotypical assumptions often hinder women from aspiring to managerial roles in organizational environments (see Heilman 2001; Koburtay et al. 2019; Castaño et al. 2019). Breaking entrenched stereotypes and challenging norms can, in turn, lead to social disapproval (Eagly & Wood 2012; Heilman 2012). Second, these can contribute to exclusion, particularly through the use of (generic) gender-marked words in job advertisements, as found in previous studies (Stout & Dasgupta 2011; Merkel 2013; Horvath & Sczesny 2016). Gender-marked language can alienate individuals

whose gender identity is not encompassed by the wording, making them feel that they are not being addressed by the advertisement. In some cases, this exclusion may be intentional, as employers may subtly indicate preferences for hiring men or women. However, in most instances, it arises from a lack of awareness. Gender-marked language thereby perpetuates stereotypes that negatively influence motivation when applying for jobs, and as a result, limit diversity in occupations. Occupational gender stereotypes can also affect salaries – for instance, language that includes women has been found to lower the estimation of salary of a professional group (Merkel 2013). That is particularly relevant as women earn less in most occupational domains (Statistikaamet 2024a) and increasing female representation of women has been found to decrease wages for both men and women (Harris 2022).

Such stereotypes also exacerbate gender inequality. Although improvements in gender equality and increased optimism about its achievement have been documented worldwide (IPSOS 2025), numerous barriers regarding gender discrimination still need to be overcome. To illustrate, 50% of respondents from 30 countries in the aforementioned study were convinced that promoting gender equality has already gone too far. Moreover, in several countries, studies have identified a growing ideological and political divide between women and men, with men's views becoming increasingly conservative and, in some cases, even overtly misogynistic (Ella Mebane et al. 2020; Burn-Murdoch 2024). This phenomenon has also been documented in Estonia. For instance, a survey (Ümarik & Tinn 2023) found that about half of male eighth grade respondents believe that men are better suited for political leadership, while nearly a quarter think that women should not be involved in politics. These attitudes underscore deep embeddedness of gender inequality in social structures, which arises from multiple sources, including language. Gendered language has been found to influence opinions about gender equality (Pérez & Tavits 2018), shown to reinforce gender asymmetry in the division of labour (see Mavisakalyan 2015) and to contribute to disparities in labour market outcomes (see van der Velde et al. 2015). Thus, these studies highlight the growing obstacles to gender equality. While language is a significant perpetuator of gender inequalities, it is also a powerful tool for challenging and combating them (Cameron 2024: 19). Consequently, addressing these issues is becoming increasingly urgent.

These results demonstrate that while Estonian is grammatically genderless, it still expresses stereotypes, overt and covert sexism and bias (Hellinger & Bußmann 2015). Based on the comparative study of gender perception in Russian and Estonian, I hypothesize that explicit gender markings used in grammatically genderless languages may often exhibit as much – or even more – bias than grammatical gender features in gendered languages. Undoubtedly, Russian and Estonian differ significantly in their linguistic structures, are spoken in distinct social environments and cultures, thereby influencing speakers' cognitive factors differently and making comparison of their gender perception somewhat challenging. While many occupations were perceived as neutral in Estonian, grammatical gender in Russian led to biased perceptions of the same occupations,

indicating a stronger tendency toward neutrality and perhaps even equality in Estonian. This could also be supported by a study of Pérez & Tavits (2018), who found that Russian-Estonian bilinguals exhibit greater openness on questions related to gender equality and rely less on gender stereotypes when interviewed in Estonian than in Russian. This is a persuasive conclusion, since in languages with grammatical gender, the gender-based categorization of surrounding concepts and phenomena is so ingrained that speakers' thinking is more heavily influenced by gendered concepts and, consequently, stereotypes (Boroditsky et al. 2003). However, certain aspects should be considered here. The gendered vocabulary in Estonian is the result of linguistic loans. Similarly, the neutralization of Estonian *mees* 'man' probably derives from Indo-European and/or Finnish influences (Karlsson 1974, cited through Puna 2006: 12). Hence, in languages like Estonian, where gender is not that frequently expressed or that deeply embedded in the language structure as it is in grammatical gender languages, any artificially introduced lexical gender marking might appear more explicit. Moreover, while these gendered terms are relatively new, gendered connotations of generic words persist despite their neutral usage, and they are less naturalized. Generic masculinity, adopted from other languages, has become entrenched in Estonian and has shaped speakers' interpretations of both the language itself and the surrounding social order of men and women. Nonetheless, most Estonian occupational titles are gender neutral in form, therefore Estonian provides more opportunities for avoiding gender-specific expression. This flexibility makes Estonian potentially more inclusive for individuals whose identities do not fall within the male–female binary.

All in all, the term 'genderless' may not be appropriate for referring to languages such as Estonian or Finnish, which lack gender in their grammatical system. This can create a false impression that these languages are somehow free from gender differentiation and discrimination, leading to the subject of gender representation being overlooked in research on such languages. This thesis aims to demonstrate that gender is indeed present in such languages, albeit expressed through different linguistic processes. The term 'grammatically genderless' would thus be more accurate. However, as the term 'gender' itself in the context of grammatical gender has sparked debates about its accuracy, and 'natural gender' has been pointed out for referring to the gender binary (Hellinger & Bußmann 2015; Bradley 2023), I propose that it would be more reasonable in general to distinguish between noun classes (i.e., grammatical gender) and the expression of gender and gender identities in language.

To sum up with the words of Cameron (2024: 71):

“(Language is) an important source of cultural information, from which we learn what ideas, feelings, beliefs and attitudes are intelligible and acceptable in the societies we belong to. Our ideas about gender are bound to be influenced by (among other things) the language in which gender is represented and discussed. [...] (Language is) an active shaper rather than just a passive reflector of our ideas about the world.”

These results demonstrate that language is far from neutral. Instead, it is a shadow of the beliefs of its speakers and social circumstances in which it is used. Through language, perceptions of gender are constantly created and reestablished, both overtly and covertly. Based on these results, a more considerate, attentive, and symmetrical use of gender markings could be recommended, on the one hand, to ensure that no gender holds a more standard or natural status over others, and to prevent language from reinforcing outdated, harmful, and socially divisive stereotypes. On the other hand, in gender expression, it would be valuable to reflect on whether differentiating gender is relevant in a given context and to consider whether more neutral expressions could be used instead. Rather than advocating for the regulation or elimination of words, I suggest fostering greater awareness of language use and its impact on individuals.

## 5.6. Limitations and further research

There are some limitations to the approaches chosen in these studies that should be addressed. These stem mainly from the particular corpora used and the research methods adopted. Although combining qualitative and quantitative analysis (the former involved in P2 and P3) does enhance the reliability of the results and conclusions (Baker & Levon 2015), both of these methods have their inherent disadvantages as well.

The study in P1, which is based solely on sports news corpora, is limited to illustrating language patterns within this specific genre. While sports news is an important part of Estonian public discourse, as previously mentioned, the findings may reflect the representations of genders in the vocabulary in a narrower context. However, these results are supported by the comparison with the Estonian Reference Corpus, which includes texts from various domains beyond sports news and yields results similar to those of the modern sports news corpora. Incorporating data from additional corpora from later periods and analyzing texts from other fields, such as politics or culture, would perhaps strengthen these findings. Furthermore, while analyzing frequencies and semantic categories of gender-marked words effectively reveals overt sexism and stereotyping in language, it may not capture more subtle patterns of gender dynamics and covert sexism. A promising avenue for future research would be a qualitative analysis of how male and female athletes (or representatives of other fields) are represented, focusing on deeper patterns of portrayal in language alongside explicit ones, and providing more nuances to the analysis of gender in language (similarly to Kane 1996; Laine 2016; Kaukonen et al. 2024a). A more qualitative analysis would enable the examination of more discursive patterns in the text. While this dissertation has a stronger emphasis on corpus linguistics, incorporating more discourse analysis in the future would complement the results, provide important insights and uncover nuances in the asymmetry of gender representation that may currently remain hidden.

Similarly, the results of the study of *tädi*- and *onu*-compounds in P3 may be limited by analyzing only the specific uses of these words – namely, five most frequent compounds referring to occupations. While providing valuable insights, this study may not fully represent all possible cases or contexts in which these compounds occur. Therefore, including more examples would further support the findings. Moreover, as the study involved a qualitative analysis of *tädi* and *onu*, by categorizing example sentences based on their context, one limitation of this approach is subjectivity, as the examination and interpretation of concordance lines often accommodate this (see Chapter 3.1.2). To build on the findings presented in this thesis, further research could employ survey methodologies to explore language users' attitudes toward *tädi* and *onu*. Such a survey could more specifically address whether these terms carry negative or derogatory connotations for language users, comparable to the results of concordance analysis.

Some instances of generic masculinity have been feminized in Estonian, as discussed in Chapter 5.2. However, a more comprehensive examination of other feminine compounds beyond the case study of *esinaine* would provide a more systematic overview of the processes of de-stereotyping in the Estonian language. In addition to this, exploring potential gender-neutral counterparts would offer deeper insights into the scope of feminist language planning. Moreover, making conclusions solely based on the frequency changes can be somewhat problematic due to the lack of accompanying discussions. Hence, survey methodologies could additionally offer more insights into the opportunities and challenges of this activism in Estonian (see Oksanen 2019), as also mentioned in Section 5.2. Feminized forms have low usage frequencies, and not all words of generic masculinity have a neutral counterpart. Surveys could better help determine which approach – feminization or neutralization – is more suitable in Estonian, according to language users, assess the acceptability of certain gender-neutral words among language users, and examine how demographic factors, such as political views and exposure to feminist discourses, might influence opinions about and attitudes toward feminist language planning. This aspect could also be explored in further experimental studies on gender perception as well (see Hammond-Thrasher & Järvikivi 2023).

This brings me to potential ideas of further experiments for exploring gender perception among Estonian language users, which could be pursued in future research. In addition to the central tendency bias, the limitations of Likert scale surveys also include explicitness. Specifically, while this method reveals overt gender biases in language, it may not uncover deeper, implicit patterns. Another limitation is that the perceptions of linguistic units were evaluated without randomly assigning participants to different groups. Therefore, this study could benefit from additional experimental methods that investigate implicit biases and involve different groups of participants. A common experimental design used to examine similar hypotheses, which I also plan to use in the future, involves sentence evaluation (see Carreiras et al. 1996; Oakhill et al. 2005; Gygax et al. 2008; Misersky et al. 2019; Hammond-Thrasher & Järvikivi 2023), as well as eye-tracking or free association (see Tremmel & Wahl 2023). Data analysis for

such experiments could also employ alternative methods and models, such as decision trees.

Lastly, these studies could benefit from more cross-linguistic comparisons, in addition to Russian in P4, as this would enhance these findings even further and help corroborate the drawn conclusions about linguistic stereotypes. More specifically, a comparison with similarly structured languages, such as Finnish, would be appropriate.

## 6. CONCLUSION

This thesis studied Estonian gender-marked vocabulary, by combining two different methodological approaches: corpus linguistics and a quasi-experimental approach. The thesis consists of an overview section and four research publications, which illustrate the different methods utilized: in P1–P3, corpora representing various texts from different periods across the past hundred years (1920s–2020s) were analyzed to identify vocabulary, examine its use in contexts and to discuss these findings in light of feminist language planning. In P4, a survey method was employed, to investigate language users' gender perception of Estonian occupational titles – both gender-neutral and gender-marked – in comparison with Russian. The data was subjected to both quantitative and qualitative analyses, with the former used to identify frequencies, statistical significance and the influence of possible variables on the survey results, and the latter to analyze usage patterns and contextualize them within a broader ideological context in the corpus analysis. The primary aim of this thesis is thus to document the existing vocabulary (specifically, gender-marked compound nouns) in Estonian, by examining their frequencies in the corpora, the semantic categories they express, their usage in contexts, and how they are perceived. Additionally, a broader goal is to uncover how Estonian gender-marked vocabulary expresses stereotypes about women and men as well as the roles attributed to them. This includes examining how gender inequalities become evident through language.

**First**, the findings revealed that compounds ending with a gender marking mostly exhibit an unbalanced frequency ratio in favour of words referring to men. This was particularly evident for compounds ending with *mees* and *naine*, with *mees*-compounds occurring predominantly, as these are used generically and represent official occupational titles in Estonian. A more balanced gender ratio was represented by compounds ending with *tädi* and *onu*, which exhibited slightly higher frequencies of compounds referring to women (i.e., *tädi*-compounds). In contrast, among words beginning with gender markings, words denoting women (i.e., *nais*-compounds) had higher frequencies. **Second**, it became evident that the use of marked vocabulary has changed moderately over time. Moreover, while there are some indications of feminization in Estonian vocabulary, with several occupational titles incorporating feminine gender markings being used more frequently over time and shedding their sexist connotations, such instances remain rare. In addition, there are no notable signs of feminist language planning in Estonian. **Third**, Estonian gender-marked vocabulary predominantly marks occupations and activities, while also describing male and female individuals through age, relations, characteristics, behavior and even cultural and social identity. Within these categories, several stereotypical assumptions emerge, with compound words referring to men and women often embodying highly stereotypical attributes and portraying men and women through distinctly masculine or feminine characteristics. **Fourth**, regarding gender-marked occupational titles, distinctions emerged in the types of occupations

associated with words referring to men and women, highlighting how language reinforces occupational gender segregation. Specifically, words referring to women were predominantly linked to occupations emphasizing communal traits, whereas words referring to men were more often associated with agentic traits. Additionally, language reveals gender stereotypes resulting from this segregation and often reflects the relative prestige of certain occupations. **Fifth**, the findings suggest that the perceptions of occupational titles in Estonian are mostly influenced by social knowledge and stereotypes. However, as the results additionally indicated, these can be reinforced by language, which perpetuates and reinforces existing biases in thought. **Finally**, these findings overall exemplify how gender bias and stereotypes are apparent in a language that has commonly been considered gender neutral. The frequency distributions as well as the use of gender-marked words in Estonian reveal the *people = male* bias, with words referring to men used generically, while feminine gender remains marked and deviant. Stereotypes and traditional gender divisions are also apparent in the categories these words express, highlighting how the vocabulary exposes the social positions of men and women. Furthermore, it became evident that the instances of generic masculinity, which are deemed neutral in Estonian, might not be perceived as such. Based on these findings, I thereby propose that grammatically genderless languages may exhibit even greater bias than gendered languages, as gender in these languages represents an artificially introduced linguistic category.

This thesis makes a contribution to the study of gender representation in languages, while also addressing a significant research gap in this field. To date, gender issues have been addressed more extensively in gendered languages. Since Estonian is a grammatically genderless language, this study thereby offers new insights into how such languages incorporate gender, and provides evidence of the bias these languages convey, despite their perception as neutral. By examining an understudied language such as Estonian, I believe that this thesis offers valuable evidence that enhances the theories of gender and language, and complements existing conclusions in this field.

Issues of gender, particularly concerning equality and human rights, remain acute in the Estonian context. The gender equality index continues to lag behind the European Union average, language use has been documented to spread sexism, misogyny (Kaukonen et al. 2024c) and male-centrism, and teenagers – especially male – have been found to exhibit growing conservatism (Ümarik & Tinn 2023). Therefore, this thesis could also introduce valuable ideas for future research, as the subject of gender (as well as sexuality) requires further investigation in Estonian linguistics and is in definite need for additional studies.

## REFERENCES

- Abele, Andrea E. 2003. The Dynamics of Masculine-Agentic and Feminine-Communal Traits: Findings From a Prospective Study. *Journal of Personality and Social Psychology* 85(4). 768–776. <https://doi.org/10.1037/0022-3514.85.4.768>
- Aikhenvald, Alexandra. 2016. *How Gender Shapes the World*. Oxford: Oxford University Press
- Aitchison, Jean. 2000. *Language change. Progress or decay?* Cambridge: Cambridge University Press.
- AMSS = Ametniku soovitussõnastik. Available at: <https://arhiiv.eki.ee/dict/ametnik/>
- Ancarno, Clyde. 2020. Corpus-Assisted Discourse Studies. In Anna De Fina and Alexandra Georgakopoulou (eds.), *The Cambridge Handbook of Discourse Studies*, 165–185. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108348195.009>
- Andrade, Chittaranjan. 2020. Sample Size and its Importance in Research. *Indian Journal of Psychological Medicine* 42, [https://doi.org/10.4103/IJPSYM.IJPSYM\\_504\\_19](https://doi.org/10.4103/IJPSYM.IJPSYM_504_19)
- Baker, Paul. 2006. *Using Corpora in Discourse Analysis*. London/New York: Continuum. <https://doi.org/10.5040/9781350933996>.
- Baker, Paul 2008. *Sexed texts. Language, gender and sexuality*. London: Equinox Publishing
- Baker, Paul. 2014. *Using Corpora to Analyze Gender*. New York: Bloomsbury Publishing
- Baker, Paul and Levon, Erez. 2015. Picking the right cherries? A comparison of corpus-based and qualitative analyses of news articles about masculinity. *Discourse & Communication* 9(2). 221–236. <https://doi.org/10.1177/1750481314568542>
- Bauer, Laurie. 2006. Compound. In Keith Brown (ed.), *Encyclopedia of Language & Linguistics* (Second edition), 719–726. Amsterdam: Elsevier Ltd
- Belle, Deborah, Ashley B. Tartarilla, Mikaela Wapman, Marisa Schlieber and Andrea E. Mercurio. 2021. “I Can’t Operate, that Boy Is my Son!”: Gender Schemas and a Classic Riddle. *Sex Roles* 85(1). <https://doi.org/10.1007/s11199-020-01211-4>
- Bem, Sandra L. and Daryl J. Bem. 1973. Does Sex-biased Job Advertising “Aid and Abet” Sex Discrimination? *Journal of Applied Social Psychology* 3(1), 6–18. <https://doi.org/10.1111/j.1559-1816.1973.tb01290.x>
- Berkutova, Veronika Valerevna. 2019. Феминативы в русском языке: лингвистический аспект [Feminatives in the Russian language: linguistic aspect] 1 (45), 7–26. Available at <https://scipress.ru/philology/articles/feminativy-v-russkom-yazyke-lingvisticheskij-aspekt.html> [04.05.2025]
- Biber, Douglas, Susan Conrad and Randi Reppen. 1998. *Corpus Linguistics. Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Blaubergs, Maija S. 1980. An Analysis of Classic Arguments Against Changing Sexist Language. *Women’s Studies International Quarterly* 3. 135–147. [https://doi.org/10.1016/S0148-0685\(80\)92071-0](https://doi.org/10.1016/S0148-0685(80)92071-0)
- Blommaert, Jan. 2005. *Discourse* (Key Topics in Sociolinguistics). Cambridge: Cambridge University Press
- Bolognesi, Marianna. 2020. *Where Words Get Their Meaning. Cognitive processing and distributional modelling of word meaning in first and second language* (Converging Evidence in Language and Communication Research). Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/ceclr.23>

- Boroditsky, Lera, Lauren A. Schmidt and Webb Phillips. 2003. Sex, Syntax and Semantics. In Dedre Gentner and Susan Goldin-Meadow (eds.), *Language in mind: Advances in the Study of Language and Cognition*, 61–79. Cambridge: MIT
- Boroditsky, Lera. 2011. How Language Shapes Thought. The languages we speak affect our perceptions of the world. *Scientific American* 304(2), 62–65
- Bradley, Evan D., Julia Salkind, Ally Moore and Sofi Teitsort. 2019. Singular ‘they’ and novel pronouns: gender-neutral, nonbinary, or both? *Proceedings of the Linguistic Society of America* 4. <https://doi.org/10.3765/plsa.v4i1.4542>
- Bradley, Evan D. 2023. Linguistic prescriptivism as social prescription. The case of gender. In Joan C. Beal, Morana Lukač and Robin Straaijer (eds.), *The Routledge Handbook of Linguistic Prescriptivism*, 213–226. <https://doi.org/10.4324/9781003095125-15>
- Braun, Virginia and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brezina, Vaclav. 2018. *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316410899>
- Bucholtz, Mary. 2014. The Feminist Foundations of Language, Gender, and Sexuality Research. In Susan Ehrlich, Miriam Meyerhoff and Janet Holmes (eds.), *The Handbook of Language, Gender, and Sexuality*, 23–47. Chichester: John Wiley & Sons. <https://doi.org/10.1002/9781118584248.ch1>
- Burgess, Diana and Eugene Borgida. 1999. Who Women Are, Who Women Should Be: Descriptive and Prescriptive Gender Stereotyping in Sex Discrimination. *Psychology Public Policy and Law* 5(3), 665–692. <https://doi.org/10.1037/1076-8971.5.3.665>
- Burnett, Heather and Céline Pozniak. 2021. Political dimensions of gender inclusive writing in Parisian universities. *Journal of Sociolinguistics* 25(5), 808–831. <https://doi.org/10.1111/josl.12489>
- Burn-Murdoch, John. 2024. A new global gender divide is emerging. *Financial Times*, 26.01.2024. Available at <https://www.ft.com/content/29fd9b5c-2f35-41bf-9d4c-994db4e12998> [04.05.2025]
- Butler, Judith. 1990. *Gender trouble. Feminism and the Subversion of Identity*. New York/London: Routledge
- Cameron, Deborah 1992. *Feminism and Linguistic Theory*. London: MacMillan Press Ltd. <https://doi.org/10.1007/978-1-349-22334-3>
- Cameron, Deborah. 2012. *Verbal Hygiene*. London/New York: Routledge
- Cameron, Deborah. 2024. *Language, sexism and misogyny*. London/New York: Routledge
- Canessa-Pollard, Valentina, David Reby, Robin Banerjee, Jane Oakhill and Alan Garnham. 2022. The development of explicit occupational gender stereotypes in children: Comparing perceived gender ratios and competence beliefs. *Journal of Vocational Behavior* 134. <https://doi.org/10.1016/j.jvb.2022.103703>
- Cao, Yang Trista and Hal Daumé III. 2020. Toward Gender-Inclusive Coreference Resolution. In Dan Jurafsky, Joyce Chai, Natalie Schluter and Joel Tetreault (eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics. 4568–4595. Available at: <https://aclanthology.org/2020.acl-main.418/> [04.05.2025]
- Carreiras, Manuel, Alan Garnham, Jane Oakhill and Kate Cain. 1996. The Use of Stereotypical Gender Information in Constructing a Mental Model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology* 49(3). 639–663. <https://doi.org/10.1080/713755647>.

- Castaño, Ana M., Yolanda Fontanil and Antonio L. García-Izquierdo. 2019. "Why can't I become a manager?"—A systematic review of gender stereotypes and organizational discrimination. *International Journal of Environmental Research and Public Health* 16(10), 1813. <https://doi.org/10.3390/ijerph16101813>
- Clarke, Victoria and Virginia Braun. 2017. Thematic analysis. *The Journal of Positive Psychology* 12(3), 297–298. <https://doi.org/10.1080/17439760.2016.1262613>
- Clyne, Michael, Catrin Norrby and Jane Warren. 2009. *Language and Human Relations. Styles of Address in Contemporary Language*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511576690>
- Conrad, Susan. 2002. Corpus Linguistic Approaches for Discourse Analysis. *Annual Review of Applied Linguistics* 22, 75–95. <https://doi.org/10.1017/S0267190502000041>
- Cooper, Robert L. 1989. *Language planning and social change*. Cambridge: Cambridge University Press
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Coutinho, Antónia, Gabriele Diewald and María Muelas-Gil. 2024. Introduction – Gender and Language: Perspectives and Challenges. *Ex Aequo* 49, 11–18. <https://doi.org/10.22355/exaequo.2024.49.02>
- Croft, William and David Alan Cruse. 2003. *Cognitive Linguistics*. (Second edition). Cambridge: Cambridge University Press.
- Croft, Alyssa, Toni Schmader and Katharina Block. 2015. An Underexamined Inequality: Cultural and Psychological Barriers to Men's Engagement With Communal Roles. *Personality and Social Psychology Review* 19(4). <https://doi.org/10.1177/1088868314564789>
- Deaux, Kay and Laurie L. Lewis. 1984. Structure of Gender Stereotypes: Interrelationships Among Components and Gender Label. *Journal of Personality and Social Psychology* 46(5), 991–1004. <https://doi.org/10.1037/0022-3514.46.5.991>
- Diewald, Gabriele and Anja Steinhauer. 2020. *Handbuch geschlechtergerechte Sprache – Wie Sie angemessen und verständlich gendern* [Handbook on gender-fair language – How to use it appropriately and comprehensively]. Duden.
- van Dijk, Teun A. 1993. Principles of Critical Discourse Analysis. *Discourse and Society* 4(2), 249–283. <https://doi.org/10.1177/0957926593004002006>
- Doleschal, Ursula and Sonja Schmid. 2001. Doing gender in Russian. Structure and perspective. In Marlis Hellinger and Hadumod Bußmann (eds.), *Gender Across Languages. The Linguistic Representation of Women and Men* 1, 253–282. Amsterdam/Philadelphia: John Benjamins Publishing Company
- Douven, Igor. 2018. A Bayesian perspective on Likert scales and central tendency. *Psychonomic Bulletin & Review* 25(3), 1203–1211. <https://doi.org/10.3758/s13423-017-1344-2>
- Eagly, Alice H. 1987. *Sex differences in social behavior: A social-role interpretation*. Lawrence Erlbaum Associates, Inc
- Eagly, Alice H., Wendy Wood and Amanda B. Diekmann. 2000. Social role theory of sex differences and similarities: A current appraisal. In Thomas Eckes and Hanns M. Trautner (eds.), *The Developmental Social Psychology of Gender*, 123–174. New York: Psychology Press
- Eagly, Alice H. and Wendy Wood. 2012. *Social Role Theory*. In Paul A.M. Van Lange, Arie W. Kruglanski and E. Tory Higgins (eds.), *Handbook of Theories of Social Psychology*, 458–476. London: Sage Publications Ltd. <https://doi.org/10.4135/9781446249222.n49>

- Eagly, Alice H., Christa Nater, David Miller, Michèle Céline Kaufmann and Sabine Sczesny. 2020. Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. *American Psychologist* 75(3).  
<https://doi.org/10.1037/amp0000494>
- Eckert, Penelope and Sally McConnell-Ginet. 2003. *Language and Gender*. Cambridge: Cambridge University Press
- EIGE 2024 = Gender Equality Index. Available at: <https://eige.europa.eu/gender-equality-index/2024> [05.02.2025]
- Engelberg, Mila. 2016. *Yleispätevä mies: suomen kielen geneerinen, piilevä ja kielio-pillistuva maskuliinisuus*. Doctoral dissertation, University of Helsinki.
- Engelberg, Mila. 2018. *Miehiä ja naisihmisiä: Suomen kielen seksismi ja sen purkaminen*. Helsinki: Tasa-arvoasiain neuvottelukunta TANE
- England, Paula, Michelle Budig and Nancy Folbre. 2002. Wages of virtue: The relative pay of care work. *Social Problems* 49(4), 455–473.  
<https://doi.org/10.1525/sp.2002.49.4.455>
- EKG I = Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael and Silvi Vare. 1995. *Eesti keele grammatika I. Morfoloogia. Sõnamoodustus*. Tallinn: Eesti Keele Instituut
- Ella Mebane, Minou, Antonio Aiello and Donata Francescato. 2020. Political Gender Gap and Social Dominance Orientation. In Dhastagir Sultan Sheriff (ed.), *Psycho-Social Aspects of Human Sexuality and Ethics*. IntechOpen.  
<https://doi.org/10.5772/intechopen.92222>
- Erdocia, Iker. 2022. Language and culture wars. The far right’s struggle against gender-neutral language. *Journal of Language and Politics* 21, 847–866.  
<https://doi.org/10.1075/jlp.21050.erd>
- Ester, Jane, Lee Maripuu, Merlin Nuiamäe, Mari-Liis Sepper and Maris Vainre. 2020. Poliitikut meedias läbi sooprisma. Tallinn: Poliitikauuringute Keskus Praxis. Available at: <https://www.praxis.ee/uploads/2019/05/Naised-poliitikas-meediaanalus-2020.pdf> [21.05.2025]
- EP 2018 = Gender-neutral Language in the European Parliament. Available at: [https://www.europarl.europa.eu/cmsdata/151780/GNL\\_Guidelines\\_EN.pdf](https://www.europarl.europa.eu/cmsdata/151780/GNL_Guidelines_EN.pdf) [04.05.2025]
- Fairclough, Norman. 2001. *Language and Power* (Second edition). London: Pearson Education LTD
- Fitch, Kristine L. 1991. The interplay of linguistic universals and cultural knowledge in personal address: Colombian Madre terms. *Communication Monographs* 58(3). 254–272. <https://doi.org/10.1080/03637759109376229>
- Formanowicz, Magdalena, Sylwia Bedyńska, Aleksandra Cislak and Sabine Sczesny. 2012. Side effects of gender-fair language: How feminine job titles influence the evaluation of female applicants. *European Journal of Social Psychology*.  
<https://doi.org/10.1002/ejsp.1924>
- Foubert, Océane. 2024. *Experience with(out) a name: coinage, conventionalisation, and hypostatisation of English neologisms in contemporary feminism*. (Doctoral dissertation). Linguistics. Université de Lille; Ludwig-Maximilians Universität. Available at: <https://theses.hal.science/tel-04726046v1> [04.05.2025]
- Friedrich, Marcus C. G., Selina Gajewski, Katja Hagenberg, Christine Wenz, and Elke Heise. 2024. Does the gender asterisk (“Gendersternchen”) as a special form of gender-fair language impair comprehensibility? *Discourse Processes*,  
<https://doi.org/10.1080/0163853X.2024.2362027>

- Frye, Marilyn. 1983. Sexism. In *The Politics of Reality: Essays in Feminist Theory*, 17–40. New York: The Crossing Press
- Gabriel, Ute and Pascal M. Gygax. 2008. Can societal language amendments change gender representation? The case of Norway. *Scandinavian Journal of Psychology* 49, 451–457. <https://doi.org/10.1111/j.1467-9450.2008.00650.x>
- Gabriel, Ute, Pascal M. Gygax, Oriane Sarrasin, Alan Garnham and Jane Oakhill. 2008. Au pairs are rarely male: Norms on the gender perception of role names across English, French, and German. *Behaviour Research Methods* 40(1). 206–212. <https://doi.org/10.3758/brm.40.1.206>
- Gabriel, Ute, Dawn M. Behne and Pascal M. Gygax. 2017. Speech vs. reading comprehension: An explorative study of gender representations in Norwegian. *Journal of Cognitive Psychology* 29, 795–808. <https://doi.org/10.1080/20445911.2017.1326923>
- Gabriel, Ute, Pascal M. Gygax and Elisabeth A. Kuhn. 2018. Neutralising linguistic sexism: Promising but cumbersome? *Group Processes & Intergroup Relations* 21(5). 844–858. <https://doi.org/10.1177/1368430218771742>
- Gabriel, Ute, Jonathan Kim, Anton Öttl, Pascal M. Gygax, Lei Cui, Jukka Hyönä and Olga Nagel. 2023. Norms on the Gender Perception of Role Nouns: Gender Ratio Data for Chinese, Finnish, and Russian. *Journal of Open Psychology Data* 11(1), 2. <http://doi.org/10.5334/jopd.73>
- Garland, Ron. 1991. The Mid-Point on a Rating Scale: Is it Desirable? *Marketing Bulletin* 2, 66–70. Available at: [http://marketing-bulletin.massey.ac.nz/V2/MB\\_V2\\_N3\\_Garland.pdf](http://marketing-bulletin.massey.ac.nz/V2/MB_V2_N3_Garland.pdf). [04.05.2025]
- Garnham, Alan, Sam Doehren and Pascal M. Gygax. 2015. True gender ratios and stereotype rating norms. *Frontiers in Psychology* 6. <https://doi.org/10.3389/fpsyg.2015.01023>
- Garnham, A., and Yuri Yakovlev. 2015. The Interaction of Morphological and Stereotypical Gender Information in Russian. *Frontiers in Psychology* 16. <https://doi.org/10.3389/fpsyg.2015.01720>
- Gaucher, Danielle, Justin Friesen and Aaron C. Kay. 2011. Evidence that gendered wording in job advertisements exists and stains gender inequality. *Journal of Personality and Social Psychology* 101(1), 109–128. <https://doi.org/10.1037/a0022530>
- Gee, James Paul. 2005. *An Introduction to Discourse Analysis*. (Second edition). London/ New York: Routledge
- Germon, Jennifer E. 2009. *Gender. A Genealogy of an Idea*. Palgrave Macmillan
- Gerritsen, Marinel. 2002. Towards a more gender-fair usage in Netherlands Dutch. In Marlis Hellinger and Hadumod Bußmann (eds.), *Gender Across Languages. The Linguistic Representation of Women and Men* 2, 81–108. Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/impact.10.10ger>
- Gillings, Matthew, Gerlinde Mautner and Paul Baker. 2023. *Corpus-Assisted Discourse Studies*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781009168144>
- Gillioz, Christelle and Sandrine Zufferey. 2020. *Introduction to Experimental Linguistics*. London: ISTE Ltd.
- Graff, Agnieszka and Elżbieta Korolczuk 2022. Gender, anti-gender and right-wing populism. Recasting the debate. In Agnieszka Graff and Elżbieta Korolczuk (eds.), *Anti-Gender Politics in the Populist Movement*, 15–37. Oxford: Routledge

- Gustafsson Sendén, Marie, Emma A. Bäck and Anna Lindqvist. 2015. Introducing a gender-neutral pronoun in a natural gender language: the influence of time on attitudes and behavior. *Frontiers in Psychology (Psychology of Language)* 6. <https://doi.org/10.3389/fpsyg.2015.00893>
- Gygax, Pascal M., Ute Gabriel, Oriane Sarrasin, Alan Garnham and Jane Oakhill. 2008. Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and Cognitive Progresses* 23(3). 464–485. <https://doi.org/10.1080/01690960701702035>
- Gygax, Pascal M. and Ute Gabriel. 2008. Can a Group of Musicians be Composed of Women? Generic Interpretation of French Masculine Role Names in the Absence and Presence of Feminine Forms. *Swiss Journal of Psychology* 67(3). 141–153. <https://doi.org/10.1024/1421-0185.67.3.143>
- Gygax, Pascal M., Ute Gabriel, Oriane Sarrasin, Jane Oakhill and Alan Garnham. 2009. Some grammatical rules are more difficult than others: The case of the generic interpretation of the masculine. *European Journal of Psychology of Education* 24(2). 235–246. <https://doi.org/10.1007/BF03173014>
- Gygax, Pascal M., Alan Garnham and Sam Doehren. 2016. What Do True Gender Ratios and Stereotype Norms Really Tell Us? *Frontiers in Psychology* 7. <https://doi.org/10.3389/fpsyg.2016.01036>
- Gu, Yueguo. 1990. Politeness phenomena in modern Chinese. *Journal of Pragmatics* 14, 237–257. [https://doi.org/10.1016/0378-2166\(90\)90082-O](https://doi.org/10.1016/0378-2166(90)90082-O)
- Haines, Elisabeth L., Kay Deaux, Nicole Lofaro. 2016. The Times They Are a-Changing ... or Are They Not? A Comparison of Gender Stereotypes, 1983–2014. *Psychology of Women Quarterly* 40(3), 1–11. <https://doi.org/10.1177/0361684316634081>
- Halliday, Michael A.K. 2007. Language and Society. In Jonathan Webster (ed.), *The Collected works of M.A.K. Halliday* 10, 43–65
- Hamilton, Mykol C. 1991. Masculine bias in the attribution of personhood: People = Male, Male = People. *Psychology of Women Quarterly* 15(3), 393–402. <https://doi.org/10.1111/j.1471-6402.1991.tb00415.x>
- Hammond-Thrasher, Stephanie and Juhani Järviäivi. 2023. Rating gender stereotype violations: The effects of personality and politics. *Frontiers in Communication* 8. <https://doi.org/10.3389/fcomm.2023.1050662>
- Harris, Jorgen. 2022. Do wages fall when women enter an occupation? *Labour Economics* 74, <https://doi.org/10.1016/j.labeco.2021.102102>
- Hasselblatt, Cornelius 2015. The representation of gender in Estonian. In Marlis Hellinger and Heiko Motschenbacher (eds.), *Gender Across Languages. The linguistic representation of women and men* 4, 125–151. <https://doi.org/10.1075/impact.36.05has>
- Haugen, Einar. 1983. The Implementation of Corpus Planning: Theory and Practice. In Juan Cobarrubias and Joshua A. Fishman (eds.), *Progress in Language Planning (Contributions to the Sociology of Language [CSL]* 31), 269–290. Berlin: De Gruyter Mouton. <https://doi.org/10.1515/9783110820584.269>
- Hayakawa, Samuel I. and Alan R. Hayakawa. 2022 [1949]. *Keel mõttes ja teos*. Tallinn: Postimees Kirjastus
- Heilman, Madeline E. 2001. Description and Prescription: How Gender Stereotypes Prevent Women's Ascent Up the Organizational Ladder. *Journal of Social Issues* 57(4), 657–674. <https://doi.org/10.1111/0022-4537.00234>

- Heilman, Madeline E. and Alice H. Eagly. 2008. Gender stereotypes are alive, well, and busy producing workplace discrimination. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1(4), 393–398. <https://doi.org/10.1111/j.1754-9434.2008.00072.x>
- Heilman, Madeline E. 2012. Gender stereotypes and workplace bias. *Research in Organizational Behavior* 32, 113–135. <https://doi.org/10.1016/j.riob.2012.11.003>
- Hekanaho, Laura. 2022. A thematic analysis of attitudes towards English nonbinary pronouns. *Journal of Language and Sexuality* 11, 190–216. <https://doi.org/10.1075/jls.21025.hek>
- Hellinger, Marlis and Hadumod Bußmann. 2015. Linguistic Representation of Women and Men. In Marlis Hellinger and Heiko Motschenbacher (eds.), *Gender across languages. The linguistic representation of women and men 4*, 1–26. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Holmes, Janet. 2001. A corpus-based view of gender in New Zealand English. In Marlis Hellinger and Hadumod Bußmann (eds.), *Gender across languages. The linguistic representation of women and men 1*, 115–136. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Holmes, Janet, Robert Sigley and Agnes Terraschke. 2009. From chairman to chairwoman to chairperson. Exploring the move from sexist usages to gender neutrality. In Pam Peters, Peter Collins and Adam Smith (eds.), *Comparative studies in Australian and New Zealand English*, 181–202. Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/veaw.g39.11hol>
- Horvath, Lisa Kristina, Elisa Merkel, Anne Maass and Sabine Sczesny. 2016. Does Gender-Fair Language Pay Off? The Social Perception of Professions from a Cross-Linguistic Perspective. *Frontiers in Psychology* 6. <https://doi.org/10.3389/fpsyg.2015.02018>
- Ibbotson, Paul. 2013. The scope of usage-based theory. *Frontiers in Psychology* 4, <https://doi.org/10.3389/fpsyg.2013.00255>
- IPSOS 2025 = *International Women's Day Survey 2025*. Global Institute for Women's Leadership, King's College London. Available at: [https://www.ipsos.com/sites/default/files/ct/news/documents/2025-03/IWD%202025%20Global%20Charts%20FINAL\\_1.pdf](https://www.ipsos.com/sites/default/files/ct/news/documents/2025-03/IWD%202025%20Global%20Charts%20FINAL_1.pdf) [04.05.2025]
- Irmen, Lisa and Nadja Roßberg. 2004. Gender Markedness of Language: The Impact of Grammatical and Nonlinguistic Information on the Mental Representation of Person Information. *Journal of Language and Social Psychology* 23(3), 272–307. <https://doi.org/10.1177/0261927X04266810>
- Jespersen, Otto. 1990 [1922]. The Woman. In Deborah Cameron (ed.), *The Feminist Critique of Language. A reader*, 201–220. London/New York: Routledge
- Jung, Kiju, Sharon Shavitt, Madhu Viswanathan and Joseph M. Hilbe 2014. Female hurricanes are deadlier than male hurricanes. *Psychological and Cognitive Sciences* 111 (24), 8782–8787. <https://doi.org/10.1073/pnas.1402786111>
- Jurafsky, Daniel. 1996. Universal Tendencies in the Semantics of the Diminutive. *Language* 72(3), 533–578. <https://doi.org/10.2307/416278>
- Justiitsministeerium 2023 = *Kuritegevus Eestis 2023*. Available at <https://www.just.ee/kuritegevus2023/> [04.05.2025]
- Kane, Mary Jo. 1996. Media coverage of the post Title IX female athlete. *Duke Journal of Gender Law and Policy* 3, 95–127.
- Kasik, Reet. 2004. *Eesti keele sõnatuletus*. Tartu: Tartu Ülikooli Kirjastus

- Kasik, Reet. 2006. Feministlik lingvistika. *Oma Keel* 2, 5–12. Available at: [https://www.emakeeleselts.ee/omakeel/2006\\_2/OK\\_2006-2\\_01.pdf](https://www.emakeeleselts.ee/omakeel/2006_2/OK_2006-2_01.pdf) [04.05.2025]
- Kasik, Reet. 2015. *Sõnamoodustus*. Tartu: Tartu Ülikooli Kirjastus
- Kask, Arnold. 1955. Fr. R. Kreutzwaldi poolt tarvitusele võetud liide *-na (-nna)* eesti keeles. *Ajaloo-keeleteaduskonna töid*, 30–42. Tartu: Tartu Riiklik Ülikool
- Kapatinski, Vsevolod. 2006. Sex Associations of Russian Generics. *Journal of Slavic Linguistics* 14(1), 17–43
- Karlsson, Fred. 1974. Sukupuoliroolien kielellisistä heijastumista. *Sananjalka* 16, 24–33
- Karro, Piret. 2022. 150 aastat Eesti feminismi. *Vikerkaar* 3, 57–112
- Kaufmann, Christiane and Gerd Bohner. 2014. Masculine generics and gender-aware alternatives in Spanish. *IZGOnZeit (Onlinezeitschrift des Interdisziplinären Zentrums für Geschlechterforschung)* (3). <https://doi.org/10.4119/izgonzeit-1310>
- Kaukonen, Elisabeth. 2022. Kuidas on lood soolise tasakaaluga eesti keeles? *Feministeerium*. Available at: <https://feministeerium.ee/kuidas-on-lood-soolise-tasakaaluga-eesti-keeles/> [04.05.2025]
- Kaukonen, Elisabeth. 2023. Stereotüüpidest siplev eesti keel. *ERR*. Available at: <https://www.err.ee/1608908267/elisabeth-kaukonen-stereotüüpidest-siplev-eesti-keel> [04.05.2025]
- Kaukonen, Elisabeth, Marie Luise Meier and Raili Marling. 2024a. *Supernaised ja äri-rüütliid: nais- ja meesjuhtide meediakajastuse ja sotsiaalmeediakasutuse analüüs*. (Research report). Tallinn: Poliitikauuringute Keskus Praxis. Available at: <https://www.praxis.ee/uploads/2024/08/Praxis.%20Nais-%20ja%20meesjuhtide%20meediakajastuste%20ja%20sotsiaalmeediakasutuse%20anal%C3%BC%C3%BCs.pdf> [04.05.2025]
- Kaukonen, Elisabeth, Ahmed Sabir and Rajesh Sharma. 2024b. How Aunt-Like Are You? Exploring Gender Bias in the Genderless Estonian Language: A Case Study. *NoDaLiDa/Baltic-HLT 2025*.
- Kaukonen, Elisabeth, Kaarel Lott, Raili Marling, Maria Murumaa-Mengel, Mari-Liis Sepper, Maarja Tinn. 2024c. *Alfamehed ja lillekesed: misogüünia ja seksismi normaliseerumine Eesti meediapildis*. (Research report). Tallinn: Mõttekoda Praxis. Available at: <https://www.praxis.ee/uploads/2024/12/Raport-lillekesed.pdf> [05.02.2025]
- Kelan, Elisabeth K. 2010. Gender Logic and (Un)doing Gender at Work. *Gender, Work and Organization* 17 (2). 175–194. <https://doi:10.1111/j.1468-0432.2009.00459.x>
- Kennison, Sheila M. and Jessie L. Trofe. 2003. Comprehending Pronouns: A Role for Word-Specific Gender Stereotype Information. *Journal of Psycholinguistic Research* 32(3). 355–378. <https://doi.org/10.1023/A:1023599719948>
- Kiesling, Scott. 2007. Men, Masculinities and Language. *Language and Linguistics Compass* 1(6), 653–673. <https://doi.org/10.1111/j.1749-818X.2007.00035.x>
- Kilgarriff, Adam, Pavel Rychlý, Pavel Smrž and David Tugwell. 2004. The Sketch Engine. Geoffrey Williams and Sandra Vessier (eds.), *Proceedings of the 11th EURALEX International Congress*, 105–115. Lorient: Université de Bretagne Sud
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubiček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý and Vít Suchomel. 2014. The Sketch Engine: Ten years on. *Lexicography* 1, 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- Kim, Jonathan K., Sarah Angst, Pascal M. Gygax, Ute Gabriel and Sandrine Zufferey. 2023. The masculine bias in fully gendered languages and ways to avoid it: A study on gender neutral forms in Québec and Swiss French. *Journal of French Language Studies* 23(1). 1–26. <https://doi.org/10.1017/S095926952200014X>

- Kirey-Sitnikova, Yana. 2021. Prospects and challenges of gender neutralization in Russian. *Russian Linguistics* 45(2). 143–158. <https://doi.org/10.1007/s11185-021-09241-6>
- Kiss, Orsolya. 2022. *Forms of address in the Tatar language spoken in Finland and Estonia. Institute of Estonian and General Linguistics*. Tartu: Tartu Ülikool. (Master's thesis). Available at: [https://dspace.ut.ee/bitstream/handle/10062/76857/Kiss\\_O\\_2022\\_mag.pdf?sequence=1&isAllowed=y](https://dspace.ut.ee/bitstream/handle/10062/76857/Kiss_O_2022_mag.pdf?sequence=1&isAllowed=y) [04.05.2025]
- Klavan, Jane, Ann Veismann and Anni Jürine. 2013. Katselised meetodid tähenduse uurimisel. *Eesti ja Soome-Ugri Keeleteaduse Ajakiri* 4(1), 17–34. <https://doi.org/10.12697/jeful.2013.4.1.02>
- Kloss, Heinz. 1969. *Research possibilities on group bilingualism: a report*. Quebec: International Center for research on Bilingualism.
- Koburtay, Tamer, Jawad Syed and Radi Haloub. 2019. Congruity between the female gender role and the leader role: A literature review. *European Business Review* 31(6), 831–846. <https://doi.org/10.1108/EBR-05-2018-0095>
- Koik, Mari. 2022. Keele-elu ja nõrgemate kaitse. *Sirp*. Available at: <https://www.sirp.ee/s1-artiklid/varamu/keele-elu-ja-norgemate-kaitse/> [04.05.2025]
- Koenig, Anne M. and Alice H. Eagly. 2014. Evidence for the Social Role Theory of Stereotype Content: Observations of Groups' Roles Shape Stereotypes. *Journal of Personality and Social Psychology* 103(7). 371–392. <https://doi.org/10.1037/a0037215>
- Koivunen, Anu and Marianne Liljeström. 2003. *Võtmesõnad: 10 sammu feministliku uurimuseeni*. Tallinn: Eesti Keele Sihtasutus
- Koppel, Kristina and Jelena Kallas. 2022. Eesti keele ühendkorpuste sari 2013–2021: mahukaim eesti keele kogu. In Margit Langemets, Maria-Maren Linkgreim and Helle Metslang (eds.), *Eesti Rakenduslingvistika Ühingu Aastaraamat, 207–228*. Tallinn: Eesti Rakenduslingvistika Ühing. <https://doi.org/10.5128/ERYa18.12>
- Kramarae, Cheri and Paula A. Treichler. 1985. *A feminist dictionary*. Boston: Pandora.
- Kress, Gunther. 1989. *Linguistic Processes in Sociocultural Practice*. (Second edition). Oxford: Oxford University Press
- Krosnick, Jon A., Allyson L. Holbrook, Matthew K. Berent, Richard T. Carson, Michael W. Hanemann, Raymond J. Kopp, Robert Cameron Mitchell, Stanley Presser, Paul A. Ruud, Kerry V. Smith, Wendy R. Moody, Melanie C. Green and Michael Conaway. 2002. The Impact of “No Opinion” Response Options on Data Quality: Non-Attitude Reduction or an Invitation to Satisfice? *Public Opinion Quarterly* 66(3). 371–403. <https://doi.org/10.1086/341394>
- Kuhar, Roman and David Paternotte. 2017. “Gender ideology” in movement: Introduction. In Roman Kuhar and David Paternotte (eds.), *Anti-Gender Campaigns in Europe. Mobilizing Against Equality*, 1–22. Lanham/New York: Rowman & Littlefield
- Kusmaryono, Imam, Nila Ubaidah and Mochamad Abdul Basir. 2022. Number of Response Options, Reliability, Validity, and Potential Bias in the Use of the Likert Scale Education and Social Science Research: A Literature Review. *International Journal of Educational Methodology* 8(4). 625–637. <https://doi.org/10.12973/ijem.8.4.625>
- Kuusik, Aet. 2015. Need on ju ainult sõnad? *Feministeerium*. Available at: <https://feministeerium.ee/need-ju-ainult-sonad/> [04.05.2025]
- Kuusik, Aet. 2024. Kuidas suhtub Eesti LGBT kogukond sõnasse kväär? *Keel ja Kirjandus* 4, 315–339. <https://doi.org/10.54013/kk796a1>
- Laine, Antti. 2016. Gender Representation of Athletes in Finnish and Swedish Tabloids. *Nordicom Review* 37(2), 83–98. <https://doi.org/10.1515/nor-2016-0012>

- Lakoff, Robin. 1973. Language and Woman's Place. *Language in Society* 2(1), 45–80. Cambridge: Cambridge University Press.  
<https://doi.org/10.1017/S0047404500000051>
- La ministre de l'Éducation nationale et de la Jeunesse 2021 = Règles de féminisation dans les actes administratifs du ministère de l'Éducation nationale, de la Jeunesse et des Sports et les pratiques d'enseignement. *Le Bulletin de L'éducation Nationale, de la Jeunesse et des Sports* 18. Available at <https://www.education.gouv.fr/bo/21/Hebdo18/MENB2114203C.htm> [04.05.2025]
- Lassonde, Karla A. and Edward J. O'Brien. 2013. Occupational stereotypes: activation of male bias in a gender-neutral world. *Journal of Applied Social Psychology* 43. 387–396. <https://doi.org/10.1111/j.1559-1816.2013.01008.x>
- Le Lamer, Jeanne 2012. *Every chairman for himself? A Study of the Influence of Feminist Language Planning on Written Media Language*. Oslo: The University of Oslo (Masters' thesis). Available at <http://urn.nb.no/URN:NBN:no-32086> [04.05.2025]
- de Lauretis, Teresa 1987. *Technologies of Gender*. Indiana: Indiana University Press.
- Levanon, Asaf and David B. Grusky. 2016. The Persistence of Extreme Gender Segregation in the Twenty-first Century. *American Journal of Sociology* 122(2). 573–619. <https://doi.org/10.1086/688628>
- Lewis, Molly and Gary Lupyan. 2020. Gender stereotypes are reflected in the distributional structure of 25 languages. *Nature Human Behavior* 4, 1021–1028. <https://doi.org/10.1038/s41562-020-0918-6>
- Liben, Lynn S., Rebecca S. Bigler and Holleen R. Krogh. 2002. Language at Work: Children's Gendered Interpretations of Occupational Titles. *Child Development* 73(3), 810–828. <https://doi.org/10.1111/1467-8624.00440>
- Liddicoat, Anthony J. 2011. Feminist language planning. *Current Issues in Language Planning* 12(1), 1–7. <https://doi.org/10.1080/14664208.2011.548314>
- Likert, Rensis. 1932. A technique for the measurement of attitudes. *Archives of Psychology*, 140, 5–55.
- Lillhannus, Ruth. 2002. Open, Closed, and Locked Images: Cultural Stereotypes and the Symbolic Creation of Reality. *Journal of Intercultural Communication* 3(1). <https://doi.org/10.36923/jicc.v3i1.383>
- Lindström, Liina, Lydia Risberg and Helen Plado. 2023. Language ideologies and beliefs about language in Estonia and Estonian language planning. *Journal of Estonian and Finno-Ugric Languages* 14, 7–48. <https://doi.org/10.12697/jeful.2023.14.1.01>
- Litosseliti, Lia 2013. *Gender and Language. Theory and Practice*. Oxon/New York: Routledge.
- Lövgren, Karin. 2013. The Swedish Tant: A Marker of Female Aging. *Journal of Women & Aging* 25, 119–137.
- Mackey, Alison and Susan M. Gass. 2021. *Second Language Research. Methodology and Design* (Third edition). London/New York: Routledge
- Makarov, Ivan. 2023. Emakeele soolistamata üraskid ja meeshinge soolistatud libaapaatia. *Objektiiv*. Available at: <https://objektiiv.ee/emakeele-soolistamata-uraskid-ja-meeshinge-soolistatud-libaapaatia/> [04.05.2025]
- Malmberg, Kristi. 2005/2006. Meeste meediakujutise muutus 1996. ja 2004. aasta Eesti Päevalehe ja Postimehe nädalalõpulisade näitel. *Ariadne Lõng* 1/2, 25–34
- Malone, Helen, Honor Nicholl and Catherine Tracey. 2014. Awareness and minimisation of systematic bias in research. *British Journal of Nursing* 23(5). 279–282. <https://doi.org/10.12968/bjon.2014.23.5.279>
- Manne, Kate. 2018. *Down Girl: The Logic of Misogyny*. Oxford: Oxford University Press

- Marling, Raili. 2010. The Intimidating Other: Feminist Critical Discourse Analysis of the Representation of Feminism in Estonian Print Media. *NORA – Nordic Journal of Feminist and Gender Research* 18(1), 7–19.  
<https://doi.org/10.1080/08038741003626767>
- Marling, Raili, Mari-Liis Sepper, Elisabeth Kaukonen and Maarja Tinn. 2024. Kes ees, see mees? Soolistatud juhtimiskultuur ja selle meediakajastused Eestis. *Ariadne Lõng* 20, 65–103
- Martin, J.R. and David Rose. 2007. *Working with Discourse. Meaning Beyond the Clause* (Second edition). London: Bloomsbury.
- Mautner, Gerlinde. 2009. Corpora and Critical Discourse Analysis. In Paul Baker (ed.), *Contemporary Corpus Linguistics*, 32–46. London: Continuum
- Mavisakalyan, Astghik. 2015. Gender in Language and Gender in Employment. *Oxford Development Studies* 43(4), 403–424.  
<https://doi.org/10.1080/13600818.2015.1045857>
- Mawson, Marlene. 2006. Sportswomanship: The cultural acceptance of sport for women versus the accommodation of cultured women in sport. In Linda K. Fuller (ed.), *Sport, Rhetoric and Gender*. London: Palgrave Macmillan, 19–30.  
[https://doi.org/10.1057/9780230600751\\_2](https://doi.org/10.1057/9780230600751_2)
- McConnell-Ginet, Sally 1988. Language and Gender. In Frederick J. Newmeyer (ed.), *Language: The Socio-Cultural Context* (Linguistics: The Cambridge Survey IV), 75–99. Cambridge: Cambridge University Press
- McConnell-Ginet, Sally. 2020. *Words matter: Meaning and Power*. Cambridge: Cambridge University Press
- McEnery, Tony and Andrew Wilson. 2001. *Corpus Linguistics: An Introduction*. (Second edition). Edinburgh: Edinburgh University Press
- McEnery, Tony and Paul Baker. 2015. Introduction. In Tony McEnery and Paul Baker (eds.), *Corpora and Discourse Studies. Integrating Discourse and Corpora* (Palgrave Advances in Language and Linguistics), 1–19. London: Palgrave Macmillan.
- Merkel, Elisa. 2013. *The Two Faces of Gender-Fair Language*. Doctoral dissertation, Padova: Università degli Studi di Padova. Available at:  
[http://paduaresearch.cab.unipd.it/6119/1/Elisa\\_Merkel\\_Thesis\\_final.pdf](http://paduaresearch.cab.unipd.it/6119/1/Elisa_Merkel_Thesis_final.pdf)  
 [04.05.2025]
- Mills, Sara. 2008. *Language and Sexism*. Cambridge: Cambridge University Press
- Mills, Sara and Louise Mullany. 2011. *Language, Gender and Feminism. Theory, Methodology and Practice*. Oxon/New York: Routledge
- Misersky, Julia, Pascal M. Gyax, Pablo Canal, Ute Gabriel, Alan Garnham, Friederike Braun, Tania Chiarini, Kjellrun Englund, Adriana Hanulikova, Anton Öttl, Jana Valdova, Lisa von Stockhausen and Sabine Sczesny. 2013. Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behaviour Research Methods* 46(3). 841–847.  
<https://doi.org/10.3758/s13428-013-0409-z>
- Misersky, Julia, Asifa Majid, Tineke M. Snijders. 2019. Grammatical Gender in German Influences How Role-Nouns Are Interpreted: Evidence from ERPs. *Discourse Processes*, 56(8), 1–12. <https://doi.org/10.1080/0163853X.2018.1541382>
- Molinaro, Nicola, Jui-Ju Su and Manuel Carreiras. 2016. Stereotypes override grammar: Social knowledge in sentence comprehension. *Brain and Language* 155–156, 36–43.  
<https://doi.org/10.1016/j.bandl.2016.03.002>
- Motschenbacher, Heiko. 2022. *Linguistic dimensions of sexual normativity. Corpus-based evidence*. New York: Routledge

- Mäearu, Sirje. 2008. Kas naine on esimees või esinaine? In Maire Raadik (ed.), *Keele- nõuanne soovitab* 4, 79. Tallinn: Eesti Keele Sihtasutus
- Nadler, Joe T., Rebecca Weston and Elora C. Voyles. 2015. Stuck in the Middle: The Use and Interpretation of Mid-Points in Items on Questionnaires. *The Journal of General Psychology* 42, 71–89. <https://doi.org/10.1080/00221309.2014.994590>
- Naffine, Ngaire. 1987. *Female Crime. The Construction of Women in Criminology*. London/New York: Routledge
- Naistekas 2006 = Kui tädilik sina oled? *Naistekas* 20.06.2006. Available at: <https://naistekas.delfi.ee/artikkel/13175211/kui-tadilik-sina-oled> [04.05.2025]
- Niedzwiecki, Patricia 1993. Women and Language. *Women of Europe Supplement no 40*. EU Commission. Available at: <https://aei.pitt.edu/34010/> [04.05.2025]
- Oakhill, Jane, Alan Garnham and David Reynolds. 2005. Immediate activation of stereotypical gender information. *Memory & Cognition* 33(6). 972–983. <https://doi.org/10.3758/BF03193206>
- Oksanen, Juho. 2019. Man-compound occupational nouns and non-sexist alternatives: Cross-linguistic perspectives. University of Eastern Finland (Masters' thesis). Available at: [https://erepo.uef.fi/bitstream/handle/123456789/21756/urn\\_nbn\\_fi\\_uef-20200071.pdf?sequence=-1&isAllowed=y](https://erepo.uef.fi/bitstream/handle/123456789/21756/urn_nbn_fi_uef-20200071.pdf?sequence=-1&isAllowed=y) [04.05.2025].
- Olt, Kerli. 2004. Sooliselt markeeritud sõnavara trükiajakirjanduses. *Tekstid ja Taustad (Lingvistiline tekstianalüüs)* 3, 90–107. Tartu: Tartu Ülikooli eesti keele osakond.
- Pakosta, Liisa. 2021. Kas „esimehe” ametinimetust kande naine peaks tundma end diskrimineerituna? *Eesti Päevaleht*. Available at: <https://epl.delfi.ee/artikkel/92889947/paeva-teema-liisa-pakosta-kas-esimehe-ametinimetust-kandev-naine-peak-tundma-end-diskrimineerituna> [04.05.2025]
- Papp, Ülle-Marike. 2013. Tervist, härra esinaine! *Delfi*. Available at: <https://www.delfi.ee/artikkel/65820186/ulle-marike-papp-tervist-harra-esinaine> [04.05.2025]
- Parent-Lamache, Annick and Yves Hallée. 2023. Exploring the effects of predominantly female jobs on demands and resources at work and consequently on health and performance in the Province of Québec, Canada. *International Archives of Occupational and Environmental Health* 96(9), 1267–1281. <https://doi.org/10.1007/s00420-023-02005-3>
- Partington, Alan, Alison Duguid and Charlotte Taylor. 2013. *Patterns and Meanings in Discourse. Theory and Practice in Corpus Assisted Discourse Studies (CADS)* (Studies in Corpus Linguistics). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Partington, Alan. 2023. Book review: Gillings, M., Mautner, G., & Baker, P. (2023). *Corpus-Assisted Discourse Studies (Cambridge Elements)*. *Journal of Corpora and Discourse Studies* 6(1), 53–60. <https://doi.org/10.18573/jcads.109>
- Pauwels, Anne. 1999. Feminist language planning: Has it been worthwhile? *Linguistik Online* 2(1). <https://doi.org/10.13092/lo.2.1043>
- Pauwels, Anne 2003. Linguistic sexism and feminist linguistic activism. In Janet Holmes and Miriam Meyerhoff (eds.), *The Handbook of Language and Gender*, 550–570. Oxford: Blackwell Publishing Ltd, <https://doi.org/10.1002/9780470756942.ch24>
- Penelope, Julia. 1988. Prescribed Passivity: The Language of Sexism. *A Feminist Ethic for Social Science Research (A Nebraska Sociological Feminist Collective)*, 119–139. New York: Edwin Mellen Press.
- Pérez, Efrén O. and Margit Tavits. 2018. Language Influences Public Attitudes toward Gender Equality. *The Journal of Politics* 81(1). <https://doi.org/10.1086/700004>

- Phillips, Webb and Lera Boroditsky. 2003. Can Quirks of Grammar Affect the Way You Think? Grammatical Gender and Object Concepts? *Proceedings of the Annual Meeting of the Cognitive Science Society* 25. 938–933.
- Piir, Milvi Martina. 2010. Sooperspektiiv ajalooõpikutes: metodoloogilisi tähelepanekuid. *Ariadne Lõng* 1–2, 142–149
- Piits, Liisi. 2015. *Sagedamate inimest tähistavate sõnade kollokatsioonid eesti keeles*. (Dissertationes Linguisticae Universitatis Tartuensis). Tartu: Tartu Ülikooli Kirjastus. Available at: <https://dspace.ut.ee/server/api/core/bitstreams/f6128ce1-4203-4cb5-8671-24f6591f4c5c/content> [05.02.2025]
- Pilvre, Barbi. 2009. *Naiste meediarepresentatsioon Eesti ajakirjanduskultuuri ja ühiskonna kontekstis*. (Dissertationes De Mediis Et Communicationibus Universitatis Tartuensis). Tartu: Tartu Ülikooli Kirjastus. Available at: <https://dspace.ut.ee/bitstreams/f945a9ab-e848-4fbb-a224-d884abdd269d/content> [05.02.2025]
- Prewitt-Freilino, Jennifer L., Andrew T. Caswell, and Emmi K. Laakso. 2012. The Gendering of Language: A Comparison of Gender Equality in Countries with Gendered, Natural Gender, and Genderless Languages. *Sex Roles* 66(3–4), 268–281. <https://doi.org/10.1007/s11199-011-0083-5>
- Pyykkönen, Pirita, Jukka Hyönä and Roger P.G. van Gompel. 2010. Activating Gender Stereotypes During Online Spoken Language Processing. Evidence From Visual World Eye Tracking. *Experimental Psychology* 57, 126–133. <https://doi.org/10.1027/1618-3169/a000016>
- Puna, Kerli. 2006. *Soospetsiifilised isikunimetused sõnaraamatutes ja tekstides*. (Master's thesis) Tartu: Tartu Ülikool. Available at: <https://dspace.ut.ee/bitstream/handle/10062/662/punakerli.pdf;jsessionid=7340E8CB8ACDF94696D863CB7CAAEEEDF?sequence=5> [04.05.2025]
- Põhjala, Priit 2013. Soolisest võrdõiguslikkusest keeles. *Postimees*. Available at <https://arvamus.postimees.ee/1110300/priit-pohjala-soolisest-vord-oiguslikkusest-keeles> [04.05.2025]
- Põldsaar, Raili. 2001. Kriitiline diskursuseanalüüs ja naisekäsitlus eesti meedias. *Ariadne Lõng* 1/2, 99–106
- Põldsam, Rebeka. 2024. Homophobic Discourses and their Soviet History in Estonia. *Folklore: Electronic Journal of Folklore* 92, 49–72. <http://doi.org/10.7592/FEJF2024.92.poldsam>
- Raadik, Lydia. 2016. *Sugude kujutamine „Eesti õigekeelsussõnaraamatu ÕS 2013” näidetes*. (Master's thesis). Tallinn: Tallinna Ülikool. Available at: <https://www.etera.ee/zoom/28488/view?page=1&p=separate> [05.02.2025]
- Rahandusministeerium 2024 = Personali- ja palgastatistika. (Palgaandmed 2018–2024). Available at: <https://fin.ee/personali-ja-palgastatistika#ametnike-palgaandmed> [04.05.2025]
- R Core Team. (2023). The R Project for Statistical Computing. <https://www.r-project.org/> [05.02.2025]
- Raudsaar, Mart. 2023. Keeleseksismist emakeelepäeval. *Elu24*. Available at: <https://elu24.postimees.ee/7732574/mart-raudsaar-keeleseksismist-emakeelepaeval> [04.05.2025]
- Redl, Theresa, Stefan L. Frank, Peter de Swart and Helen de Hoop. 2021. The male bias of a generically-intended masculine pronoun: Evidence from eye-tracking and sentence evaluation. *PLOS ONE*. <https://doi.org/10.1371/journal.pone.0249309>

- Renström, Emma A., Anna Lindqvist and Marie Gustafsson Sendén. 2022. Are Gender-Neutral Pronouns Really Neutral? Testing a Male Bias in the Grammatical Genderless Languages Turkish and Finnish. *Journal of Language and Social Psychology* 42. <https://doi.org/10.1177/0261927X221146229>
- Renström, Emma A. 2025. The implementation of neo- and nonbinary pronouns: a review of current research and future challenges. *Frontiers in Psychology* 15. <https://doi.org/10.3389/fpsyg.2024.1507858>
- Ricciardelli, Rosemary, Kimberley A. Clow and Philip White. 2010. Investigating Hegemonic Masculinity: Portrayals of Masculinity in Men's Lifestyle Magazines. *Sex Roles* 63, 64–78. <https://doi.org/10.1007/s11199-010-9764-8>
- Romaine, Suzanne. 2001. A corpus-based view of gender in British and American English. In Marlis Hellinger and Hadumod Bußmann (eds.), *Gender Across Languages. The Linguistic Representation of Women and Men* 1, 153–176. <https://doi.org/10.1075/impact.9.12rom>
- Ross, Kristiina. 1996. Naine eesti keele grammatikas. *Vikerkaar* 11–12, 102–106
- Schneider, Joseph W. and Sally L. Hacker. 1973. Sex Role Imagery and Use of the Generic “Man” in Introductory Texts: A Case in the Sociology of Sociology. *The American Sociologist* 8(1), 12–18
- Schneider, Angela 2000. On the definition of ‘woman’ in the sport context. In Torbjörn Tännsjö and Claudio Tamburrini (eds.), *Values in Sport: Elitism, Nationalism, Gender Equality and the Scientific Manufacturing of Winners*, 123–138. London/New York: Taylor & Francis Group
- Schulz, Muriel R. 1990 [1975]. The semantic derogation of woman. In Cameron, Deborah (ed.), *The Feminist Critique of Language: a Reader*, 134–147. London/New York: Routledge
- Silveira, Jeanette. 1980. Generic Masculine Words and Thinking. *Women's Studies International Quarterly* 3(2–3), 165–178. [https://doi.org/10.1016/S0148-0685\(80\)92113-2](https://doi.org/10.1016/S0148-0685(80)92113-2)
- Sinclair, John McHardy 2004. *Trust the Text. Language, corpus and discourse*. Roland Carter (ed). London, New York: Routledge
- Slovská, Iryna. 2022. Що таке сексизм? (What is sexism?) *Аналітично-порівняльне правознавство* 2, 96–100. <https://doi.org/10.24144/2788-6018.2022.02.17>
- Spender, Dale. 1990. *Man Made Language* (Second edition). London: Pandora Press
- Szuba, Agnieszka, Theresa Redl and Helen de Hoop. 2022. Are Second Person Masculine Generics Easier to Process for Men than for women? Evidence from Polish. *Journal of Psycholinguistic Research* 51, 819–845. <https://doi.org/10.1007/s10936-022-09859-7>
- Stahlberg, Dagmar, Sabine Sczesny and Friederike Braun. 2001. Name Your Favorite Musician. Effects of Masculine Generics and of Their Alternatives in German. *Journal of Language and Social Psychology* 20(4). 464–469. <https://doi.org/10.1177/0261927X01020004004>
- Stahlberg, Dagmar, Friederike Braun, Lisa Irmen, Sabine Sczesny. 2007. Representation of the Sexes in Language. In Klaus Fiedler (ed.), *Social Communication*. (Frontiers of Social Psychology), 163–187.
- Statistikaamet 2022 = Rahvastiku majanduslik aktiivsus, elatusallikad ja tööränne. Available at: <https://rahvaloendus.ee/et/tulemused/rahvastiku-majanduslik-aktiivsus-elatusallikad-ja-tooranne> [05.02.2025]
- Statistikaamet 2024a = Sooline palgalõhe on aastaga vähenenud. Available at: <https://stat.ee/et/uudised/sooline-palgalohe-aastaga-vahenenud> [04.05.2025]

- Statistikaamet 2024b = Palgarakendus. Available at <https://palgad.stat.ee/> [04.05.2025]
- Stefanowitsch, Anatol. 2020. *Corpus Linguistics. A guide to the methodology*. Berlin: Language Science Press.
- Stout, Jane G. and Nilanjana Dasgupta. 2011. When He Doesn't Mean You: Gender-Exclusive Language as Ostracism. *Personality and Social Psychology Bulletin* 36(6). 757–769. <https://doi.org/10.1177/0146167211406434>
- Stubbs, Michael 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell
- Sullivan, Gail M. and Anthony R Artino Jr. 2013. Analyzing and Interpreting Data From Likert-Type Scales. *Journal of Graduate Medical Education* 5(4), 541–542. <https://doi.org/10.4300/JGME-5-4-18>
- Zamfir, Ionel. 2024. Gender inequality in sport. *At a glance* (European Parliament Think Tank).
- Tainio, Liisa. 2006. Gender in Finnish Language Use: Equal, Inequal and/or Queer? *WEB-FU: Wiener elektronische Beiträge des Instituts für Finno-Ugristik*, 1–19. Available at: <https://webfu.univie.ac.at/texte/tainio.pdf> [04.05.2025]
- Talbot, Mary. 1998. *Language and Gender. An Introduction*. Malden: Blackwell Publishers Inc.
- Tavast, Arvi. 2022. Milleks meile eesti keel? *ERR*, <https://www.err.ee/1608593491/arvi-tavast-milleks-meile-eesti-keel> [04.05.2025]
- Taylor, Charlotte and Frazer Heritage. 2024. Introduction. In Frazer Heritage and Charlotte Taylor (eds.), *Analysing Representation. A Corpus and Discourse Textbook*. London: Routledge
- Thomas, Linda and Shân Wareing 1999. *Language, Society and Power: an Introduction*. London: Psychology Press. <https://doi.org/10.4324/9780203426968>
- Tibblin, Julia, Joost van der Weijer, Jonas Granfeldt and Pascal M. Gygax. 2023. There are more women in joggeur:euses than in joggeurs: On the effects of gender-fair forms on perceived gender ratios in French role nouns. *Journal of French Language Studies* 33, 28–51 <https://doi.org/10.1017/S0959269522000217>
- Tinits, Peeter 2020. Eesti Rahvusraamatukogu digilabori tööriistad tekstimeterjali ligipääsuks ja töötlemiseks. *Zenodo*. <https://doi.org/10.5281/zenodo.3953795>
- Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Tremmel, Manuela and Ingrid Wahl. 2023. Gender stereotypes in leadership: Analyzing the content and evaluation of stereotypes about typical, male, and female leaders. *Frontiers in Psychology* 14, <https://doi.org/10.3389/fpsyg.2023.1034258>
- Toots, Leena Karin. 2023. *LGBT+ kogukonnaga seotud sõnavara Eesti LGBT+ aktivistide taskuhäälingutes*. Tartu: Tartu Ülikool (Bachelor's thesis). Available at: <https://dspace.ut.ee/server/api/core/bitstreams/87141158-2302-413b-8718-fd8c6a966928/content> [05.02.2025]
- Tudisco, Jordan J. 2021. Queering the French Académie: Reclaiming linguistic authority for trans and non-binary people. *Toronto Working Papers in Linguistics*, 43(1). <https://doi.org/10.33137/twpl.v43i1.35952>
- Uued Uudised 2023 = Marufeminist ründab emakeelepäeval „seksistlikku” eesti keelt. Available at: <https://uueduudised.ee/marufeminist-rundab-emakeelepaeval-seksistlikku-eesti-keelt/> [04.05.2025]
- Vaidya, Shubhangi. 2019. Gender Stereotypes. In Walter Leal Filho, Annabela Marisa Azul, Luciana Brandli, Amanda Lange Salvia, and Tony Wall (eds.), *Gender Equality*, 1–9. [https://doi.org/10.1007/978-3-319-70060-1\\_27-1](https://doi.org/10.1007/978-3-319-70060-1_27-1)

- van der Velde, Lucas, Joanna Tyrowicz and Joanna Siwinska. 2015. Language and (the estimates of) the gender wage gap. *Economics Letters* 136, 165–170. <https://doi.org/10.1016/j.econlet.2015.08.014>
- Vervecken, Dries, Bettina Hannover and Ilka Wolter. 2013. Changing (S)expectations: How gender fair job descriptions impact children's perceptions and interest regarding traditionally male occupations. *Journal of Vocational Behaviour* 82(3), 208–220. <https://doi.org/10.1016/j.jvb.2013.01.008>
- Viik, Kadi. 2015. Mis on sugu? 2. osa. *Feministeerium*. Available at: <https://feministeerium.ee/mis-on-sugu-vol-2/> [05.02.2025]
- West, Candace and Don H. Zimmermann. 1987. Doing Gender. *Gender and Society* 1(2), 125–151. <https://doi.org/10.1177/0891243287001002002>
- Wodak, Ruth. 1989. Introduction. In Wodak, Ruth (ed.), *Language, Power and Ideology. Studies in Political Discourse*, 13–20. Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/ct.7>
- Wodak, Ruth. 2015. *The Politics of Fear: What Ring-Wing Populist Discourses Mean*. London: Sage Publications Ltd
- Wolfson, Nessa and James Manes. 1980. The compliment as a social strategy. *Research on Language and Social Interaction* 13(3), 391–410. <http://doi.org/10.1080/08351818009370503>
- Xiao, Hualin, Brent Strickland and Sharon Peperkamp. 2023. How Fair is Gender-Fair Language? Insights from Gender Ratio Estimations in French. *Journal of Language and Social Psychology* 48(1), 82–106. <https://doi.org/10.1177/0261927X221084643>
- ÕS 1999 = *Eesti keele sõnaraamat ÕS 1999*. Tiiu Erelt (ed.) Tiiu Erelt, Tiina Leemets, Sirje Mäearu, Maire Raadik. Tallinn: Eesti Keele Sihtasutus.
- ÕS 2006 = *Eesti õigekeelsussõnaraamat ÕS 2006*. Tiiu Erelt (ed.) Tiiu Erelt, Tiina Leemets, Sirje Mäearu, Maire Raadik. Eesti Keele Instituut. Tallinn: Eesti Keele Sihtasutus.
- ÕS 2013 = *Eesti õigekeelsussõnaraamat ÕS 2013*. Maire Raadik (ed.) Tiiu Erelt, Tiina Leemets, Sirje Mäearu, Maire Raadik. Tallinn: Eesti Keele Sihtasutus.
- ÕS 2018 = *Eesti õigekeelsussõnaraamat ÕS 2018*. Maire Raadik (ed.). Tiiu Erelt, Tiina Leemets, Sirje Mäearu, Maire Raadik. Eesti Keele Instituut. Tallinn: EKSA.
- Ümarik, Meril and Maarja Tinn. 2023. ICCS 2022. aasta uuringu Eesti aruande põhisõnumid. In Meril Ümarik and Maarja Tinn (eds.), *Noorte Hääl Demokraatia Kriisis: Kodanikuhariduse Väljakutsed. IEA 2022. aasta rahvusvahelise kodanikuhariduse uuringu (ICCS 2022) Eesti tulemused*, 7–14. Tallinn: Tallinna Ülikool
- ÜS 2025 = *Eesti Keele Instituudi ühendsõnastik 2025*. Eesti Keele Instituut, Sõnaveeb. Available at: <https://sonaveeb.ee>. [05.02.2025]

## SUMMARY IN ESTONIAN

### Sugu soota keeles.

#### Eesti keele sooliselt markeeritud sõnavara ja selle tajumine

Väitekiri vaatleb eesti keele sooliselt markeeritud sõnavara, täpsemalt sõnavara kasutussagedusi ja -mustreid, tähendusrühmi ning keelekasutajate tõlgendust nimetatud sõnadest. Töö teoreetilise raamistiku moodustab feministlik lingvistika ning uurimisprobleemile on lähenetud kahe eri meetodi kaudu. Esiteks on töös kasutatud korpuslingvistilisi meetodeid, eesmärgiga kaardistada olemasolevat sõnavara ja analüüsida selle kasutust kontekstides, ning teiseks kvaasiekspriimentaalset meetodit, et vaadelda, kuidas tajuvad teatud sooliselt markeeritud sõnu keelekasutajad. Sooliselt markeeritud sõnadena vaadeldakse selles väitekirjas liitsõnu, mis sisaldavad soolist tähendust väljendavat komponenti (nt *esimees, jahinaine*).

**Feministlik lingvistika** ühendab keeleteaduse soouuringutega, keskendudes sugude kohta käiva sotsiaalse teabe avaldumisele keeles. Feministlik lingvistika tugineb hüpoteesil, et keel ei ole neutraalne, vaid selles peegelduvad sugude kohta käivad ühiskondlikud hoiakud, normid ja arusaamad, mida keel omakorda kujundab ja taastoodab (Cameron 1992; Talbot 1998; Pauwels 2003). Sarnane põhimõte kehtib ka kriitilises diskursuseanalüüsis, mille kohaselt ei saa keelt käsitleda ühiskonnast eraldi, vaid ühiskondlikud ja kultuurilised aspektid ning keelekasutus mõjutavad pidevalt üksteist vastastikku (Kress 1989; Fairclough 2001; Blommaert 2005; Halliday 2007). Ka sugu on üks selline sotsiaalne aspekt, mille ehitamises ja kajastamises keel niisiis osaleb ning mis omakorda mõjutab keelekasutusmustreid. Ehkki soolisi identiteete on palju, keskendutakse selles väitekirjas naistele ja meestele, kuna eestikeelses sõnavaras on nende sugude väljendamiseks selgelt eristatavad markerid.

Feministlikus lingvistikas nähakse keelt ka sugudevaheliste võimuhete ja patriarhaalse ühiskonnakorra peegeldajana. Ajalooliselt ebaühtlaselt jaotunud võimuvaherkord, kus mehed on olnud paremal positsioonil, domineerinud mitmes valdkonnas, sealjuures keelekorralduses ja -poliitikas, on kaasa toonud mehi normina kohtleva keele (Spender 1990; Cameron 1992: 4; Pauwels 1999; Pauwels 2003). Lisaks lähtutakse feministlikus lingvistikas eeldusest, et keeles väljenduvad **soostereotüübid** ehk üldistava ja hoiakulise iseloomuga uskumused sugude, nende omaduste ja rollide kohta ning **seksism** ehk negatiivne või diskrimineeriv suhtumine indiviidi tema soo alusel (Frye 1983; Heilman 2001; Manne 2018; Vaidya 2019). Viimasest on tuletatud mõiste **seksistlik keel**, mis tähistab seksismi ilmnamist keeles ja millel eristatakse kaht vormi: otsene (või keelesüsteemi sisene) ning kaudne (keelesüsteemi väline) seksism (Mills 2008: 11, 34). Esimene tähistab oma olemuselt diskrimineerivat keelt. Näiteks esineb paljudes keeltes, sh eesti keeles **üldistatud maskuliinsust**, mille puhul kasutatakse just meessugu tähistavat keelendit ka sugudeüleses tähenduses, viitamaks mis tahes soost inimesele (Cooper 1989; Pauwels 2003; Stahlberg jt 2007: 169; Hellinger & Bußmann

2015: 9). Sellised keelendid tugevdavad feministliku kriitika järgi arusaama, et „universaalset inimest” esindab mees, samas kui naised ja muude sooliste identiteetide esindajad on teisejärgulised, ebastandardised ja seetõttu tihti markeeritud staatuses (Koivunen, Liljeström 2003: 11). Teine seksistliku keele vorm – kaudne seksism – tähistab kontekstipõhist soolist diskrimineerimist keelekasutuses. Kaudne seksism on varjatud ning see on sageli peidetud huumori või iroonia taha. Keelelise seksismi ja soolise kallutatuse vähendamise nimel on tegutsenud **feministlik keelearendus**. See on keelelise aktivismi vorm, mille eesmärk on parendada ja tasakaalustada eri soost indiviidide esindatust keeles; pakkuda välja lahendusi soolise ebavõrdsuse vähendamiseks; suurendada teadlikkust soolisest kallutatusest ning muuta keelt seeläbi võrdõiguslikumaks (Pauwels 1999; Pauwels 2003; Liddicoat 2011). Feministliku keelearenduse kaks põhilist strateegiat on keele feminiseerimine ehk naissoole viitavate vormide loomine ja kasutamine ning neutraliseerimine ehk sooneutraalsete keelendite eelistamine (Gabriel jt 2018). Feministlikku keelearendust on alustatud mitmes keelekogukonnas, nagu hollandi, inglise, vene jne (vt Romaine 2001; Gerritsen 2002; Diewald & Steinhauer 2020, viidatud Friedrich jt 2024 kaudu; Kirey-Sitnikova 2021).

Sellest johtuvalt püüab siinne väitekiri välja selgitada, mil moel annab eesti keele sooliselt markeeritud sõnavara edasi ühiskondlikke hoiakuid naiste ja meeste ning neile omistatud rollide kohta, kas ja kuidas tulevad sõnavara kaudu esile soostereotüübid ja sooline ebavõrdsus ning kas ja kuidas ühtivad saadud tulemused feministliku keelearenduse põhimõtetega. Kuna keele ja soo uurimise väljund on loomulikustunud keelekasutusmuustrite küsimärgistamine ja kriitiline analüüsimine (Cameron 1992; Eckert & McConnell-Ginet 1995: 9), vaatleb väitekiri ka seda, kuidas on sooliselt markeeritud sõnavara kasutusmuustrid eesti keeles tavapärastunud ning mis on selle taga. Niisiis on väitekirjal laiemalt kaks eesmärki: esiteks dokumenteerida olemasolevat sõnavara sugude nimetamiseks ja teiseks analüüsida, mida see sõnavara sugude ühiskondlike positsioonide kohta ütleb. Väitekirjas püütakse vastata järgmistele uurimisküsimustele:

1. Millised sooliselt markeeritud isikule viitavad nimisõnad esinevad eesti keeles ning milline on naistele ja meestele osutavate sõnade sageduste vaherkord?
2. Kas sooliselt markeeritud sõnade sagedused ja kasutusmuustrid on aja jooksul muutunud? Mil määral peegeldavad need muutused arenevaid ühiskondlikke norme sugude kohta ja ühtivad feministliku keelearenduse põhimõtetega?
3. Milliseid tähendusrühmi eesti keele sooliselt markeeritud isikule viitavad nimisõnad väljendavad?
4. Milliseid ameteid tähistavad sooliselt markeeritud ametinimetused ning kas selliste sõnade kasutus peegeldab naiste ja meeste positsioone ja rolle ühiskonnas?

5. Kuidas keelekasutajad eestikeelseid (sooliselt markeeritud) ametinimetusi tajuvad ja kas seda arusaama mõjutab keel või ühiskondlikud stereotüübid? Kuidas see erineb vene keele kõnelejate ametinimetuste tõlgendamisest?
6. Kas eesti keele sooliselt markeeritud sõnavara kasutamine paljastab naiste ja meeste positsioonid ühiskonnas ning kuidas annab see ühtekokku edasi soostereotüüpe ja soolist kallutatust?

Väitekirja koosneb katuspeatükist ja neljast teaduspublikatsioonist, mis on järjestatud vastavalt nende kirjutamisele ja avaldamisele. Samuti peegeldab publikatsioonide järjestus uurimuse etappe: esimesed kolm publikatsiooni esindavad korpuslingvistilisi meetodeid, et saada sõnavarast ülevaade, ning neljas publikatsioon analüüsib keelelisi stereotüüpe keelekasutajate vaatenurgast kvaasiekspriimendi kaudu. Katuspeatükk tutvustab artikleid siduvat teoreetilist raamistikku ja uurimuste tegemisel läbitud metodoloogilisi samme, võtab kokku tulemused ja lõpetuseks esitab materjali uurimusteks tulevikus.

Tüpoloogiliselt liigitub eesti keel soota keelte hulka. See tähendab, et eesti keeles ei ole soolisi nimisõnaklasse ega soolise tähendusega asesõnu, vaid sugu väljendatakse ainult leksikaalsete vahenditega (vt Stahlberg jt 2007; Hellinger & Bußmann 2015). Seetõttu on eesti keelt peetud märkimisväärselt sooneutraalsemaks kui grammatilise sooga keeli ning selle tulemusel ei ole eesti keele soolist kallutatust – sarnaselt teiste soota keeltega – ka nii põhjalikult uuritud. Samuti napib feministliku fookusega uurimusi Eesti keeleteaduse maastikul – nimetada võiks näiteks Puna (2006), Raadiku (2016) ja Piitsa (2015) uurimistööd. Väitekirja annab niisiis olulise panuse uurimislünga täitmisse nii kohalikul kui ka rahvusvahelisel tasandil, täiendades materjali keele soolise kallutatuse kohta. Teema on vajalik ja aktuaalne, sest see puudutab inimõiguste küsimust laiemalt. 2024. aasta Eesti soolise võrdõiguslikkuse indeks jääb 10 punktiga alla Euroopa Liidu keskmisele (EIGE 2024), lisaks on viimasel aastakümnel avaldatud uuringuid, mis on tõendanud misogüünia ja konservatiivsuse levikut Eesti noorte (meeste) seas (Ümarik & Tinn 2023; Kaukonen jt 2024c). Niisiis on soolise võrdõiguslikkusega seonduvad küsimused Eesti kontekstis jätkuvalt päevakajalised.

Väitekirja aluseks oleva materjali moodustavad esmalt tekstid keelekorpustest ehk suurtest elektroonilistest tekstiandmete kogudest (vt Biber jt 1998; Tognini-Bonelli 2001; Conrad 2002; Stefanowitsch 2020). Täpsemalt on kasutatud 2020. aasta Delfi ja ERR-i spordiudiste korpust, mis on autori koostatud (P1); eesti keele koondkorpust 1990–2008 (P1 ja P2); Eesti Rahvusraamatukogu digitaalarhiivi DIGAR eestikeelseid väljaandeid, milles on esindatud tekstid aastatest 1920–1979 (P2) ning eesti keele 2021. aasta ühendkorpuse (vt Koppel & Kallas 2022) aastatel 2013 ja 2021 koostatud veebitekstide allkorpust (P2 ja P3). Analüüs publikatsioonis P1 keskendub spetsiifilisemalt spordiudiste keelekasutusele ning spordiudiste korpuse võrdlusele mahukama ja heterogeensema koondkorpusega. Kõikidele korpustele peale DIGARi on ligi pääsetud korpuspääringusüsteemiga SketchEngine (Kilgarriff jt 2004; 2014); tekstid DIGARi

materjalidest koguti abivahendiga (Tinits 2020). Korpustest tehti päring sõna-loendi (ingl *wordlist*) funktsiooni kasutades, et tuvastada otsitava soomarkeeringuga lõppevate (*-mees, -naine, -poiss, -tüdruk, -tädi, -onu*) ja algavate (*nais-, mees-*) sõnade sagedusloendid. Kuna P2 sisaldab sõna *esinaine* analüüsi eesti-keelsetes tekstides sajandi jooksul, tehti selles publikatsioonis DIGARi arhiividest päring lemmaga *esinaine*. Sageduste puhul vaadeldi nii sõnesagedusi ehk absoluutsagedusi kui ka tüübisagedusi ehk eri lemmade arve (Stefanowitsch 2020: 311–313). Tulemused laaditi üksikasjalikuma analüüsi tegemiseks alla Exceli failidena; tähendusrühmade vaatlemisel kategoriseeriti sõnu liitsõna täiendosa alusel. Samuti on publikatsioonides P1–P3 kasutatud konkordantsi funktsiooni (ingl *concordance*), mis kuvab otsitavat keeleüksust ümbritsevas kontekstis ja võimaldab niisiis analüüsida seda sisaldavaid lausenäiteid. Nii publikatsioonides P2 kui ka P3 moodustati lausenäidete analüüsimiseks 100-lause juhuvalimid.

Teiseks koosnes materjal keelekasutajatelt veebiküsitluse kaudu kogutud andmetest (publikatsioon P4). Eesti- ja venekeelsete ametinimetuste analüüsimiseks kasutati kaht 5-punktilisel Likerti skaalal (Likert 1932) põhinevat küsimustikku, mis oli täitmiseks üleval LimeSurvey platvormil. Mõlemas keeles sisaldas küsitlus ametinimetusi (36 eesti ja 34 vene küsitluses), mis esindavad erinevaid soo väljendamise struktuure (sooneutraalsed vs. soomarkeeringuga ametinimetused eesti keeles ning grammatiliselt meessoost vs. naissoost ametinimetused vene keeles). Samuti eristus valitud ametites jaotus naiste ja meeste vahel: küsitlustes on ametid, kus on ülekaalus nii naised kui ka mehed, ning ametid, milles on naiste ja meeste protsent tasakaalus. Küsimused imiteerisid töökuulutust ning vastajate ülesanne oli ette kujutada, et esitatud ametikohtadele otsivad töötajaid eri ettevõtted (nt *Lux Express AS otsib bussijuhti*). Vastajad pidid tegema otsuse, kas igale positsioonile eelistatakse tööle võtta *ainult mehi, pigem mehi, pole vahet, pigem naisi* või *ainult naisi*. Kokku oli 581 vastajat eesti ja 326 vastajat vene keele küsitluses. Seesugune disain liigitub kvaasiekspérimentaalse lähenemise alla, sest kui sarnaselt eksperimentaalsete uuringutega on eesmärk analüüsida keeleüksuste tajumist ja töötlemist ning tuvastada potentsiaalsete tegurite mõju sellele, siis kvaasiekspérimentendis puudub osalejate juhuslik jaotamine kontroll- ja katserühmadesse (Mackey & Gass 2021: 269). Nii oli ka nendes küsitlustes, sest kõik vastajad hindasid kõiki ametinimetusi, mis esitati neile juhuslikus järjekorras.

Korpusuurimustes (publikatsioonid P1–P3) on meetodina rakendatud **korpuslingvistikal põhineva diskursuseanalüüsi** printsiipe. Selle meetodi puhul vaadeldakse, kuidas edastavad tekstid teavet ideoloogiate, võimuhete ja sotsiaalsete tõekspidamiste kohta (vt Wodak 1989; van Dijk 1993; Fairclough 2001). Materjal tekstide kriitiliseks analüüsiks pärineb aga korpustest, mis võimaldavad keelemustreid tuvastada tarkvara abil suures hulgas representatiivsetes tekstides (Conrad 2002; Partington jt 2013; McEnery & Baker 2015). Siinne doktoritöö asetub selle meetodi raamistikus rohkem korpuslingvistika poolele, kasutades diskursuseanalüüsi põhimõtteid pigem teoreetiliste lähtekohtadena, et vaadelda korpustest kogutud keelemustrite taga olevaid laiemad sotsiaalseid tähendusi.

Analüüsis ühendatakse kvantitatiivne ja kvalitatiivne lähenemine, keskendudes ühelt poolt keelenähtuste sagedustele ja eri muutujate mõjule tõlgendustes ning teiselt poolt tulemusi seletavale laiemale ideoloogilisele kontekstile. Nende kahe meetodi kombineerimine aitab järelduste ja tõlgenduste usaldusväärsust suurendada ning annab nüansirikkama pildi teatud keelemustritest (Baker & Levon 2015: 233; Partington jt 2013: 10–12; McEnery & Baker 2015: 2). Statistiline analüüs ja andmete visualiseerimine on tehtud tarkvaraga R (R Core Team 2023).

Esiteks selgus analüüsi tulemustest, et soomarkeeringuga lõppevate sõnade hulgas on meessooliitega sõnadel nii sõne- kui ka tüübisagedustes suur ülekaal – näiteks spordisõnavara puhul on *naine*-lõpuliste liitsõnade sõnesagedused palju väiksemad, moodustades *mees*- ja *naine*-lõpulistest sõnadest kokku vaid 5% Delfi ja ERRi korpuses ning 2% koondkorpuses (**uurimisküsimus 1**). Meestele osutavate sõnade ülekaal on märgatav ka 2021. aasta veebikorpuse *poiss*- (72%) ja *tüdruk*-lõpulist (28%) liitsõnade sõnesagedustes. Sageduste poolest kõige tasakaalukam sooliitega lõppevate sõnade rühm on aga *tädi*- ja *onu*-lõpulisel liitsõnad, mille puhul esines naisele osutavaid sõnu pisut rohkem (52% *tädi*-lõpulisel vs. 48% *onu*-lõpulisel liitsõnu). Seevastu soomarkeeringuga algavate sõnade puhul on osakaalud võrdsemad ning naissoomarkeeringuga sõnu esineb pisut ülekaalukamalt – nende sõnesagedused moodustavad sooliitega algavatest sõnadest näiteks 60% Delfi ja ERRi korpuses ning 35% koondkorpuses. Samuti näitas analüüs, et 10 000 sõna kohta on sooliitega lõppevatest sõnadest kõige sagedamini kasutatud *mees*-lõpulisel liitsõnu ning seda ülekaalukalt Delfi ja ERRi korpuses. Meessoomarkeeringuga lõppevate sõnade ülekaalu põhjendab nende kasutamine sugudeüleselt ehk üldistatult, samuti esineb just näiteks *mees*-lõpp paljudes ametlikes ametinimetustes. Samuti on mehele osutavad sõnad tugevamalt leksikaliseerunud ja seega suure tõenäosusega tajutakse neid vähem sugu rõhutavana. *Tädi*-lõpulistel liitsõnade väike ülekaal võib aga olla tingitud nende laiemast kasutusalaast, produktiivsemast esinemisest sõnamoodustuses võrreldes teiste naissoomarkeeringutega ning sellest, et *tädi* ja *onu* esindavad kuuluvad rohkem argikeelsesesse kasutusse.

Eri ajaperioode esindavate korpuste (1990–2008, 2020 ja 2021) sooliselt markeeritud sõnade sageduste võrdlus vihjab, et sõnavara kasutamine on aja jooksul vaid õige pisut muutunud (**uurimisküsimus 2**). Näiteks 2020. aasta Delfi ja ERRi korpuses on *naine*-lõpulistel liitsõnade sõnesageduste osakaal veidi suurem kui 1990–2008 aastate tekstide sisaldavas koondkorpuses. Samas aga näib, et nii *mees*-lõpulistel kui ka *mees*-algulistel liitsõnade kasutus on ajapikku sagenenud, sest nende sagedus 10 000 sõna kohta on suurem uuemates tekstides. Nende andmete põhjal ei saa aga feministliku keelearenduse mõju kohta eesti keeles täpseid järeldusi teha, sest esiteks ei ole selleteemalisi diskussioone kuigi palju olnud ning teiseks vähemalt spordisõnavara sageduste näitel paistab, et normi esindab jätkuvalt mees. Teisalt leidub märke eesti keele feminiseerumisest, mõningate *naine*-lõpulistel liitsõnade kasutuse muutumisest, nende ebstandarduse, erandlikkuse ja kallutatuse vähenemisest. Seda illustreerib sõna *esinaine*: kuni 2022. aastani kehtis keelekorraldusallikates soovitus kasutada *esinaist* ainult

naisorganisatsiooni juhi nimetamiseks (Mäearu 2008). Selgus, et sellises kitsendatud tähenduses esines see sõna küll ülekaalukalt 1920–1949. aastate tekstides, kuid tänapäevases keelekasutuses on *esinaine* sagedamini ka mis tahes organisatsiooni juht. Sellist kasutuse muutust põhjendab asjaolu, et sada aastat tagasi ei olnud naistel võrreldes tänapäevaga võimalik mis tahes asutuste juhina töötada, seega taastoodab kirjeldatud soovitus seksistlikke arusaamu, et naised saavad olla juhid vaid naistega seotud valdkondades. Soovitus muuteti 2022. aastal EKI ühendsõnastikus (ÜS 2025). Samuti on ajapikku pisut suurenenud mõningate *naine*-lõpuliste sõnade, nagu *jahinaine* ja *turvanaine*, kasutussagedused.

Kõige enam väljendavad eesti keele sooliselt markeeritud liitsõnad 1) ametid, tegevusi ja hobisid, aga ka 2) vanust, 3) suhteid ja sugulust, 4) kultuurilist ja sotsiaalset identiteedi ehk rahvust, rassi, religiooni ja seksuaalsust ning 5) välimust ja iseloomu (**uurimisküsimus 3**). Sooliselt markeeritud sõnad spordivaldkonnas kirjeldavad kõige sagedamini teatud spordiala (*rallimees*, *naisjooksja*), mistõttu võib need samuti ametit ja tegevust väljendavate sõnade sekka paigutada. Lisaks ilmnevad sõnavaras ka elutute nähtuste, tegelaste ja olendite (*tuulepoiss*, *vässuonu*, *bondionu*, *haldjatädi*) ning nimede ja kasutajanimede (*Rita-tädi*, *Vasaraonu*) tähendusrühmad. Huvitav tulemus on, et elutuid nähtusi tähistavad sõnad sisaldavad sagedamini meesoomarkeeringut *-onu* või *-poiss*. Sugudele viitav sõnavara on kohati üsna stereotüüpne, kirjeldades naisi ja mehi neile traditsiooniliselt omistatud tunnuste kaudu – näiteks mehi seostatakse autodega ja naisi (hoolitsetud) välimusega, kusjuures naiste välimust kirjeldavad sõnad on tihti negatiivsed (*notsutädi* ehk sea välimusega *naine*, *pandatädi* ehk tumedate silmaalustega *naine*).

Sealjuures ametit ja tegevust väljendavad sõnad tõendavad soolist segregatsiooni Eesti tööturul (**uurimisküsimus 4**). Ametinimetused on feministliku kriitika oluline fookus, sest ametivaldkonnas on sugudevahelised võimusuhted iseäranis ebavõrdselt jagunenud ning ametid on seega tugevalt seotud soostereotüüpide tekkimisega (Deaux & Lewis 1984; Penelope 1988; Baker 2014). Sellele viitavad näiteks eesti keele *tädi*- ja *onu*-lõpulised ametid tähistavad liitsõnad, mis ilmestavad traditsioonilisi soorolle ning seda, milliseid ametid seostatakse rohkem naiste ja milliseid meestega. *Tädi*-lõpp esineb sagedamini klienditeenindust, puhastusteenindust ning haridust ja sotsiaaltööd väljendavate ametite nimetustes (nagu *garderoobitädi*, *koristajatädi*, *õpetajatädi*), samal ajal *onu*-lõpp on rohkem seotud korrakaitse, ehituse, äri ja teadusega (*nagu politseionu*, *toruonu*, *teadlaseonu*). Selline jaotus on ilmne ka statistikas, mille järgi esineb naisi ülekaalukalt sotsiaalhoolekandes ja hariduses, samal ajal kui mehed domineerivad ehituses (Statistikaamet 2022). Lisaks selgus publikatsioonis P3 käsitletud *tädi*- ja *onu*-lõpuliste liitsõnade kasutuse analüüsist, et selliseid lastekeelt iseloomustavaid sõnu kasutatakse sageli ka iroonilises või suisa negatiivses kontekstis, sõnaga nimetatud isiku naeruvääristamiseks, mõnitamiseks või halvustamiseks.

Niisiis ilmestab sõnavara ühelt poolt sotsiaalset reaalsust ning teiselt poolt kinnistab ja tugevdab stereotüüpe, mis tulenevad nii selle reaalsuse pinnalt – isiklikest tähelepanekutest selle kohta, millise soo esindajat sagedamini mingis

rollis nähakse – kui ka eelarvamustest, ühiskondlikest ettekirjutustest ning allikatest nagu meedia (Eagly 1987; Eagly jt 2000; Eagly & Wood 2012; Gygax jt 2016). On leitud, et ka keel tugevdab stereotüüpe ja äratav soolist kallutatust keelekasutajate mõtlemises juba mõne keeleüksuse töötlemise hetkel (Kennison & Trofe 2003; Oakhill jt 2005; Gygax jt 2008; Gygax & Gabriel 2008; Misersky jt 2019). Sarnaseid tendentse ilmestavad ka tulemused eesti keele kohta (**uurimisküsimus 5**). Publikatsioonis P4 kajastatud küsitlusest selgus, et valdavat osa eesti keele sooneutraalsetest ametinimetustest tajuvad ka keelekasutajad neutraalselt. Samas tekkisid kallutatud tõlgendused ametite puhul, mida peavad ka reaalse statistika järgi rohkem kas naised või mehed (nagu küünetehnik, sõdur). Kallutatust näib tugevdavat ka keelekasutus, kuna eesti keele sooliselt markeeritud ametinimetuste puhul esines rohkem kallutatud vastuseid ning üldistatud meessooliituga ametinimetusi (*esimees, turvamees*) tõlgendati rohkem mehele osutavatena kui nende sooneutraalseid vasteid (*juht, turvatöötaja*). Kooskõlas varasemate uurimustega (Kapatinski 2006; Garnham & Yakovlev 2015) näitasid tulemused ka seda, et vene keele kõnelejad kalduvad rohkem toetuma sotsiaalsele teadmistele, mis on keelelisest teabest sageli tugevam – stereotüüpselt naiselikke või mehelikke ameteid tähistavaid sõnu tõlgendatakse vastavalt, olenevata sellest, millise grammatilise soo vormis on vastav sõna. Tulemused peakomponentide analüüsist ja logistilisest regressioonist näitasid samuti, et mõlema keele puhul mõjutavad kallutatud tõlgendust kõige rohkem sotsiaalsed teadmised ja keelelised vormid.

Nende tulemuste põhjal võib teha mitmesuguseid järeldusi selle kohta, kuidas sooneutraalseks peetud eesti keel peegeldab soostereotüüpe, keelelist seksismi ja soolist kallutatust (**uurimisküsimus 6**). Esiteks paistab sooliselt markeeritud sõnade kasutussagedustest, et mehi tajutakse sagedamini universaalse inimesena: näiteks meessoomarkeeringut esineb nende üldistatud kasutuse tõttu ülekaalukamalt ning seda kasutatakse rohkem ka elutute nähtuste isikustamiseks. Samuti esineb *mees*-lõpp produktiivsemalt sõnamoodustuses, sest materjalis leidis lisaks leksikaliseerunud *mees*-lõpuliste sõnadele ka markantsemaid uusmoodustisi. Spordiuudiste puhul on kirjutaja neid sõnu tõenäoliselt kasutanud eesmärgiga tähistada sportlast üldiselt. Lisaks kuna enamik spordiajakirjanikke on mehed, võivad mehed üldistatud maskuliinsust kasutada sagedamini, sest on leitud, et mehed tõlgendavad seda kaasavamalt, tajudes, et sellised sõnad viitavad neile endile (Hamilton 1991; Redl jt 2021). Niisiis leidub ka eesti keeles üldistatud maskuliinsust, millele on feministlikus kriitikas tähelepanu osutatud ning mille põhjal esindab inimest vaikimisi mees. Sealjuures ehkki *mees*-lõpuliste sõnade naisele osutavad vasted on aja jooksul pisut rohkem kasutust leidnud ja nende kasutus paistab olevat ajakohastatud, on need eesti keeles veel erandlikud ja ebatavalised, nagu on näha nende väikestest kasutussagedustest. Teisalt aga näib, et naissoost tegijale viidatakse eesti keeles sagedamini täiendosaga *nais*-, sest see võib võrreldes naissoomarkeeringuga liitsõna põhiosas tunduda neutraalsem ja formaalsem (vrd *naisportlane/spordinaine*). Samas kuna soomarkeering liitsõna täiendosas tähistab ebatraditsioonilisele ametile siirdunud isikut, vihjab *nais*-

alguliste liitsõnade sagedasem kasutus naisele kui teatud kõrvalekaldele normist (Romaine 2001: 158; Puna 2006: 75).

Teiseks avalduvad eesti keele sooliselt markeeritud ametit tähistavatest sõnadest sooline segregatsioon, traditsioonilised soorollid ja ühiskondlikud normid. Naisi nähakse sagedamini rollides, mis seostuvad empaatilisuse, hoolivuse, laste kasvatamise ja kogukondlikkusega, samas kui mehi rollides, mida iseloomustavad aktiivsus, domineerimine, enesekehtestamine ja agentsus (Heilman 2001). Sõnavara peegeldab samal ajal ka seda, milliseid rolle ja ameteid vastupidi ei peeta iseloomulikuks naistele või meestele – naisi ei nähta nii sageli ehituses või korrakaitstes, mehi aga hariduses ja sotsiaaltöös. Ka muid tähendusrühmi esindav sõnavara kipub naisi ja mehi kujutama stereotüüpsete omaduste kaudu. Kolmandaks, sotsiaalne reaalsus, keel, keelekasutajate kognitiivsed protsessid ja stereotüübid on omavahel ühise tervikuna toimiv kooslus, nii et kõik osad selles koosluses mõjutavad üksteist vastastikku. Keeles avalduvad ühiskonnas kehtivaid tõekspidamised ning sealhulgas viimistleb keel seda, millisena eri soo esindajaid nähakse, sest keel ei ole pelgalt maailma „passiivne peegeldaja”, vaid „aktiivne kujundaja” (Cameron 2024: 71). Sotsiaalne reaalsus viib keele soolise kallutatuseni, mis omakorda viib keelekasutajate mõtlemises tekkiva kallutatuseni, mis omakorda viib stereotüüpide tekkimiseni. Ka eesti keele kasutajate mõtlemises tekitavad kallutatust keelelised vihjed, ent need paistavad töötavat paralleelselt stereotüüpide ja sotsiaalse reaalsusega. Seda ilmestab näiteks asjaolu, et publikatsioonis P4 tehtud küsitluse kallutatud vastused tulenesid soolisest markeerimusest, kuid need markeerimused tähistavad enamasti vastavalt kas traditsiooniliselt mehelikke või naiselikke ameteid, nagu *jahimees* või *ämmaemand*. Niisiis leidub märke sellest, et sugusid markeeriv keel toimib stereotüüpide tugevdajana ja seda ka sootute keelte puhul. Õigupoolest on võimalik, et see esineb sootute keelte sooliselt markeeritud sõnavara puhul veel jõulisemalt kui grammatilise sooga keeltes, sest seesuguste keeltes ei esinda soomarkeerimused niivõrd keelestruktuuri sisse ehitatud osa kui suhteliselt uusi keelelisi laene. Näiteks meessoomarkeerimused ja üldistatud maskuliinsus eesti keele ametinimetustes, nagu *esimees*, on kasutusele võetud tõenäoliselt indoeuroopa ja/või soome keele eeskujul (Karlsson 1974, viidatud Puna 2006: 12 kaudu; ÜS 2025). Seega kuna sellistes keeltes on soo väljendamiseks kasutusele võetud markerid vähem juurdunud ja loomulikustunud kui grammatilise sooga keeltes, siis selle tulemusel on nende soolised konnotatsioonid ka vähem taandunud ning need rõhutavad üldistatud kasutusest hoolimata sugu eksplitsiitsemalt. Sellest hoolimata sisaldavad soota keeled nagu ka eesti keel laialdasemalt võimalusi erisuguste sooliste identiteetide kaasamiseks ja nimetamiseks ning seeläbi on sellistel keeltel ka suurem tõenäosus olla sooliselt võrdõiguslikum.

Väitekirjas kasutatud lähenemistel on ka mõningad kitsaskohad, samuti pakub väitekirja palju alust edasiste uurimuste jaoks. Esiteks, publikatsioonis P1 esitatud analüüs põhineb spordivaldkonda illustreerival sõnavaral, mistõttu võivad selles esitatud tulemused näitlikustada pigem kitsama konteksti kui üldkeele kasutustreid. Samas on ka spordiuudised oluline osa Eesti avalikust diskursusest ja keelekasutajate mõttemaailma kujundajatest. Teiseks, sooliselt markeeritud sõnade

sageduste ja tähendusrühmade analüüs paljastab küll eksplitsiitsed stereotüübid, kuid ei pruugi olla abiks varjatute soolistatuse muustrite tuvastamisel. Niisiis võiks seesugust uurimust täiendada pisut kvalitatiivsema sugude kujutamise analüüsiga (vt Marling jt 2024; Kaukonen jt 2024a). Kolmandaks, kasutusmuustrite analüüsis ja igasuguses tähenduslikus kategoriseerimises on kitsaskohaks ka subjektiivsus, sest keelenähtuste tõlgendamine lähtub uurija sisekaemusest. Seda ei ole alati võimalik vältida, kuid seda saab vähendada näiteks uurija kui emakeelse kõneleja keeletunnetuse ning korpuse pakutavate suurte andmete abil (Baker 2006: 18; Mautner 2009: 45; Klavan jt 2013: 18; Stefanowitsch 2020: 15). Ning neljandaks, publikatsioonis P4 käsitletud küsitluses kippusid vastused kalduma skaala keskpunkti. See on aga möödapääsmatu nähtus, mis kaasneb tihti seesugustel skaaladel põhinevate küsitlustega (vt Malone jt 2014; Douven 2018; Kusmaryono jt 2022). Samuti on küsitluse üks kitsaskoht kvaasiekspérimentaalne ülesehitus, mistõttu võiks ka seda laiendada täiendavate jätku-uuringutega, mis oleksid eksperimentaalse lähenemisega. Näiteks on selliste hüpoteeside ja uurimisprobleemi vaatlemiseks sageli kasutatud katsedisaini, mis põhineb ette antud lausete vastuvõetavuse hindamisel (vt Carreiras jt 1996; Oakhill jt 2005; Gyax jt 2008; Misersky jt 2019; Hammond-Thrasher & Järvikivi 2023). Täiendavate eksperimentaalsete meetodite abil saaks tuvastada ka keelekasutajate mõtlemises tekkivaid kaudsemaid stereotüüpe.

Kokkuvõtteks, keskendudes soolise kallutatuse uurimisele soota keeles, mis esindab alakäsitletud fookust keele ja soo teemalises uurimustes, täidab siinne väitekiri olulise uurimislünga ja täiendab teema kohta olemasolevat kirjandust. Erisuguste mehhanismidega keelte analüüsimine kõnealuses võtmes on oluline ka seetõttu, et see annab uut teavet soo väljendamise dünaamikate kohta keeles, laiendab keeleliste stereotüüpide kohta kehtivaid teooriaid ning pakub seeläbi põhjalikumat ja terviklikumat arusaama sellest, kuidas sugu kui sotsiaalne nähtus toimib. Seda ajastul, mil arusaamad sugude kohta on ühelt poolt mitmekesisumas, kuid teiselt poolt seisavad silmitsi vastupanuga traditsiooniliste normide (taas)kinnistamise näol.

## **PUBLICATIONS**

## CURRICULUM VITAE

**Name:** Elisabeth Kaukonen  
**Date of Birth:** 04.07.1995  
**Citizenship:** Estonian  
**Address:** Jakobi 2–431, 51005 Tartu  
**E-mail:** elisabeth.kaukonen@ut.ee

### Education:

2021–2025 University of Tartu, Estonian and Finno-Ugric Linguistics (PhD)  
2023 November Radboud University, Centre for Language Studies, visiting scholar  
2019–2021 University of Tartu, Estonian and Finno-Ugric Linguistics (MA)  
2014–2017 Tallinn University, Estonian philology (BA)  
2008–2014 Rocca al Mare School  
2004–2008 Tallinn Co-educational Gymnasium  
2002–2004 Kuusalu High School

### Professional experience:

2023–2025 University of Tartu, Institute of Estonian and General Linguistics, Junior Research Fellow in Estonian and Finno-Ugric Linguistics  
2019– freelance language editor  
2018–2019 Rocca al Mare School, teacher of Estonian language and literature

### Research projects:

01.01.2023–31.03.2025 CERV, Praxis Think Tank “Power of Women Leaders” (principal investigator: Raili Marling)

### Scholarly publications:

Kaukonen, Elisabeth, Polina Oskolskaia, Liina Lindström & Raili Marling. 2025. Gender, language and labour: gender perception of Estonian and Russian occupational titles. *Frontiers in Communication* 9. <https://doi.org/10.3389/fcomm.2024.1454022>

Marling, Raili, Mari-Liis Sepper, Elisabeth Kaukonen & Maarja Tinn. 2024. Kes ees, see mees? Soolistatud juhtimiskultuur ja selle meediakajastused Eestis. *Ariadne Lõng* 20, 65–103

Kaukonen, Elisabeth. 2023. Cleaning aunts and police uncles in action. Unveiling gender dynamics in Estonian compound words. *Journal of Estonian and Finno-Ugric Linguistics* 14(3), 131–171. <https://doi.org/10.12697/jeful.2023.14.3.05>

- Kaukonen, Elisabeth. 2023. Kes on esinaine? *Keel ja Kirjandus* 66 (3), 328–336. <https://doi.org/10.54013/kk783a4>
- Kaukonen, Elisabeth. 2022. Sooliselt markeeritud sõnad eesti spordiuudistes. *Keel ja Kirjandus* 6, 526–545. <https://doi.org/10.54013/kk774a3>
- Praakli, Kristiina, Mari-Liis Korkus, Aive Mandel, Elisabeth Kaukonen, Annika Kängsepp, Triin Aasa, Kristel Algvere, Helen Eriksoo, Marion Mägi, Getri Tomson, Liina Lindström. 2022. „Mis keeles ma räägin, I don’t know”. Eesti sisuloojate inglise keele kasutusest YouTube’is. *Philologia Estonica* 7, 263–291. <https://doi.org/10.22601/PET.2022.07.10>

### **Other publications:**

- Kaukonen, Elisabeth. 2025. Viisakad härrad ja väarikad prouad. Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 24.01.2025
- Kaukonen, Elisabeth. 2025. Kokteilid, mis toovad õuele. Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 10.01.2025
- Kaukonen, Elisabeth. 2024. Kalamehed ja keelearendus. *Vikerkaar* 12, 81–86
- Kaukonen, Elisabeth, Kaarel Lott, Raili Marling, Maria Murumaa-Mengel, Mari-Liis Sepper & Maarja Tinn. 2024. *Alfamehed ja lillekesed: misogüünia ja seksismi normaliseerumine Eesti meediapildis*. Research report. Tallinn: Mõttekoda Praxis
- Kaukonen, Elisabeth. 2024. Kes rämmareid mööda käib, on laadahobune. Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 15.11.2024.
- Kaukonen, Elisabeth. 2024. Elas metsas normaalne mutt. Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 01.11.2024
- Kaukonen, Elisabeth. 2024. Keeleminutid. Hommik onuga ehk kuidas keel kõneleb sugudest. *ERR Kultuur* 21.10.2024
- Kaukonen, Elisabeth. 2024. Astudes keele kujundeisse. Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 11.10.2024
- Kaukonen, Elisabeth, Marie-Luise Meier & Raili Marling. 2024. *Supernaised ja ärirüütliid: nais- ja meesjuhtide meediakajastuse ja sotsiaalmeediakasutuse analüüs*. Research report. Tallinn: Poliitikauuringute Keskus Praxis
- Kaukonen, Elisabeth. 2024. Hõi ja hoi, hoi ja hõi, raamatust ma leidsin koi. Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 28.06.2024
- Kaukonen, Elisabeth. 2024. Kes tööd teeb, see sugu ei karda. Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 31.05.2024
- Kaukonen, Elisabeth. 2024. Mis teete, karvaviilad ja koorelõuad? Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 02.02.2024
- Kaukonen, Elisabeth. 2024. Klaasike musklis veini, palun! Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 19.01.2024
- Kaukonen, Elisabeth. 2024. Lähme mimmutame mõnuga! Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 12.01.2024

- Kaukonen, Elisabeth. 2023. Kutsume mõned tibid ja kutid välja ning paneme peo püsti! Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 27.10.2023
- Kaukonen, Elisabeth. 2023. Õigekiri on vaid üks kujund keeleteaduse kindakirjas. *Novaator* 26.09.2023
- Kaukonen, Elisabeth. 2023. Ons tuul siis tüdruk või poiss? Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 16.06.2023
- Kaukonen, Elisabeth. 2023. Sugu maailmas ja keeltes. Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 14.04.2023
- Kaukonen, Elisabeth. 2023. Stereotüüpides siplev eesti keel. *ERR* 14.03.2023
- Kaukonen, Elisabeth. 2022. Naljaonu ja pahaonu ehk Mida teevad onud liitsõnades? Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 18.11.2022
- Kaukonen, Elisabeth. 2022. Kuidas on lood soolise tasakaaluga eesti keeles? *Feministarium* 01.11.2022
- Kaukonen, Elisabeth. 2022. Tädi mitu rolli. Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 29.07.2022
- Kaukonen, Elisabeth. 2022. Digimajanduses tagab edu keeleteoimetaja? *Postimees* 18.05.2022
- Risberg, Lydia & Elisabeth Kaukonen. 2022. Sugude säuts. Mees on naine on isik? Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 18.02.2022
- Kaukonen, Elisabeth. 2021. Tähelepanu! Valmis olla! Start! Ankrunaised ja mängumehed spordiajakirjanduses. *Oma Keel* 2, 3–11

## ELULOOKIRJELDUS

**Nimi:** Elisabeth Kaukonen

**Sünniaeg:** 04.07.1995

**Rahvus:** eestlane

**Address:** Jakobi 2–431, 51005 Tartu

**E-post:** elisabeth.kaukonen@ut.ee

### Haridus:

2021–2025	Tartu Ülikool, eesti ja soome-ugri keeleteadus (PhD)
2023 November	Radboudi Ülikool, Centre for Language Studies, külalis-doktorant
2019–2021	Tartu Ülikool, eesti ja soome-ugri keeleteadus (MA)
2014–2017	Tallinna Ülikool, eesti filoloogia (BA)
2008–2014	Rocca al Mare Kool
2004–2008	Tallinn Ühisgümnaasium
2002–2004	Kuusalu Keskkool

### Töökogemus:

2023–2025	Tartu Ülikool, Eesti ja Üldkeeleteaduse instituut, eesti ja soome-ugri keeleteaduse nooremteadur
2019–	vabakutseline keeleteoimetaja
2018–2019	Rocca al Mare kool, eesti keele ja kirjanduse õpetaja

### Teadusprojektid:

01.01.2023–31.03.2025 CERV, Mõttekoda Praxis „Naisjuhtide võim” (vastutav täitja: Raili Marling)

### Teaduspublikatsioonid:

Kaukonen, Elisabeth, Polina Oskolskaia, Liina Lindström & Raili Marling. 2025. Gender, language and labour: gender perception of Estonian and Russian occupational titles. *Frontiers in Communication* 9. <https://doi.org/10.3389/fcomm.2024.1454022>

Marling, Raili, Mari-Liis Sepper, Elisabeth Kaukonen & Maarja Tinn. 2024. Kes ees, see mees? Soolistatud juhtimiskultuur ja selle meediakajastused Eestis. *Ariadne Lõng* 20, 65–103

Kaukonen, Elisabeth. 2023. Cleaning aunts and police uncles in action. Unveiling gender dynamics in Estonian compound words. *Journal of Estonian and Finno-Ugric Linguistics* 14(3), 131–171. <https://doi.org/10.12697/jeful.2023.14.3.05>

- Kaukonen, Elisabeth. 2023. Kes on esinaine? *Keel ja Kirjandus* 66 (3), 328–336. <https://doi.org/10.54013/kk783a4>
- Kaukonen, Elisabeth. 2022. Sooliselt markeeritud sõnad eesti spordiuudistes. *Keel ja Kirjandus* 6, 526–545. <https://doi.org/10.54013/kk774a3>
- Praakli, Kristiina, Mari-Liis Korkus, Aive Mandel, Elisabeth Kaukonen, Annika Kängsepp, Triin Aasa, Kristel Algvere, Helen Eriksoo, Marion Mägi, Getri Tomson, Liina Lindström. 2022. „Mis keeles ma räägin, I don’t know”. Eesti sisuloojate inglise keele kasutusest YouTube’is. *Philologia Estonica* 7, 263–291. <https://doi.org/10.22601/PET.2022.07.10>

### **Muud publikatsioonid:**

- Kaukonen, Elisabeth. 2025. Viisakad härrad ja väarikad prouad. Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 24.01.2025
- Kaukonen, Elisabeth. 2025. Kokteilid, mis toovad õuele. Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 10.01.2025
- Kaukonen, Elisabeth. 2024. Kalamehed ja keelearendus. *Vikerkaar* 12, 81–86
- Kaukonen, Elisabeth, Kaarel Lott, Raili Marling, Maria Murumaa-Mengel, Mari-Liis Sepper & Maarja Tinn. 2024. *Alfamehed ja lillekesed: misogüünia ja seksismi normaliseerumine Eesti meediapildis*. Uuringuraport. Tallinn: Mõttekoda Praxis
- Kaukonen, Elisabeth. 2024. Kes rämmareid mööda käib, on laadahobune. Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 15.11.2024.
- Kaukonen, Elisabeth. 2024. Elas metsas normaalne mutt. Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 01.11.2024
- Kaukonen, Elisabeth. 2024. Keeleminutid. Hommik onuga ehk kuidas keel kõneleb sugudest. *ERR Kultuur* 21.10.2024
- Kaukonen, Elisabeth. 2024. Astudes keele kujundeisse. Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 11.10.2024
- Kaukonen, Elisabeth, Marie-Luise Meier & Raili Marling. 2024. *Supernaised ja ärirüütliid: nais- ja meesjuhtide meediakajastuse ja sotsiaalmeediakasutuse analüüs*. Uuringuraport. Tallinn: Poliitikauuringute Keskus Praxis
- Kaukonen, Elisabeth. 2024. Hõi ja hoi, hoi ja hõi, raamatust ma leidsin koi. Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 28.06.2024
- Kaukonen, Elisabeth. 2024. Kes tööd teeb, see sugu ei karda. Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 31.05.2024
- Kaukonen, Elisabeth. 2024. Mis teete, karvaviilad ja koorelõuad? Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 02.02.2024
- Kaukonen, Elisabeth. 2024. Klaasike musklis veini, palun! Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 19.01.2024
- Kaukonen, Elisabeth. 2024. Lähme mimmutame mõnuga! Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 12.01.2024

- Kaukonen, Elisabeth. 2023. Kutsume mõned tibid ja kutid välja ning paneme peo püsti! Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 27.10.2023
- Kaukonen, Elisabeth. 2023. Õigekiri on vaid üks kujund keeleteaduse kindakirjas. *Novaator* 26.09.2023
- Kaukonen, Elisabeth. 2023. Ons tuul siis tüdruk või poiss? Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 16.06.2023
- Kaukonen, Elisabeth. 2023. Sugu maailmas ja keeltes. Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 14.04.2023
- Kaukonen, Elisabeth. 2023. Stereotüüpides siplev eesti keel. *ERR* 14.03.2023
- Kaukonen, Elisabeth. 2022. Naljaonu ja pahaonu ehk Mida teevad onud liitsõnades? Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 18.11.2022
- Kaukonen, Elisabeth. 2022. Kuidas on lood soolise tasakaaluga eesti keeles? *Feministarium* 01.11.2022
- Kaukonen, Elisabeth. 2022. Tädi mitu rolli. Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 29.07.2022
- Kaukonen, Elisabeth. 2022. Digimajanduses tagab edu keeleteoimetaja? *Postimees* 18.05.2022
- Risberg, Lydia & Elisabeth Kaukonen. 2022. Sugude säuts. Mees on naine on isik? Keelesäuts (Vikerraadio). *Eesti Rahvusringhääling* 18.02.2022
- Kaukonen, Elisabeth. 2021. Tähelepanu! Valmis olla! Start! Ankrunaised ja mängumehed spordiajakirjanduses. *Oma Keel* 2, 3–11

## DISSERTATIONES PHILOLOGIAE ESTONICAE UNIVERSITATIS TARTUENSIS

1. **Ülle Viks.** Eesti keele klassifikatoorne morfoloogia. Tartu, 1994.
2. **Helmi Neetar.** Deverbaalne nominaaltuletus eesti murretes. Tartu, 1994.
3. **Ülo Valk.** Eesti rahvausu kuradi-kujutelm. Tartu, 1994.
4. **Arvo Eek.** Studies on quantity and stress in Estonian. Tartu, 1994.
5. **Reet Kasik.** Verbid ja verbaalsubstantiivid tänapäeva eesti keeles. Tartu, 1994.
6. **Silvi Vare.** Nimi- ja omadussõnatuletus tänapäeva eesti kirjakeeles. Tartu, 1994.
7. **Heiki-Jaan Kaalep.** Eesti keele ressursside loomine ja kasutamine keele- tehnoloogilises arendustöös. Tartu, 1998.
8. **Renate Pajusalu.** Deiktikud eesti keeles. Tartu, 1999.
9. **Vilja Oja.** Linguistic studies of Estonian colour terminology. Tartu, 2001.
10. **Küllü Habicht.** Eesti vanema kirjakeele leksikaalsest ja morfosüntaktilisest arengust ning Heinrich Stahli keele eripärast selle taustal. Tartu, 2001.
11. **Pire Teras.** Lõunaeesti vokaalisüsteem: Võru pikkade vokaalide kvaliteedi muutumine. Tartu, 2003.
12. **Merike Parve.** Väited lõunaeesti murretes. Tartu, 2003.
13. **Toomas Help.** Sõnakeskne keelemudel: Eesti regulaarne ja irregulaarne verb. Tartu, 2004.
14. **Heli Laanekask.** Eesti kirjakeele kujunemine ja kujundamine 16.–19. sajandil. Tartu, 2004.
15. **Peeter Päll.** Võõrnimed eestikeelses tekstis. Tartu, 2005.
16. **Liina Lindström.** Finiitverbi asend lauses. Sõnajärg ja seda mõjutavad tegurid suulises eesti keeles. Tartu, 2005.
17. **Kadri Muischnek.** Verbi ja noomeni püsiühendid eesti keeles. Tartu, 2006.
18. **Kanni Labi.** Eesti regilaulude verbisemantika. Tartu, 2006.
19. **Raili Pool.** Eesti keele teise keelena omandamise seaduspärasusi täis- ja osasihitise näitel. Tartu, 2007.
20. **Sulev Iva.** Võru kirjakeele sõnamuutmissüsteem. Tartu, 2007.
21. **Arvi Tavast.** The translator is human too: a case for instrumentalism in multilingual specialised communication. Tartu, 2008.
22. **Evar Saar.** Võrumaa kohanimed analüüs enamlevinud nimeosade põhjal ja traditsioonilise kogukonna nimesüsteem. Tartu, 2008.
23. **Pille Penjam.** Eesti kirjakeele *da-* ja *ma-*infinitiiviga konstruktsioonid. Tartu, 2008.
24. **Kristiina Praakli.** Esimese põlvkonna Soome eestlaste kakskeelne keelekasutus ja koodikopeerimine. Tartu, 2009.
25. **Mari Mets.** Suhtlusvõrgustikud reaajas: võru kõnekeele varieerumine kahes Võrumaa külas. Tartu, 2010.
26. **Karen Kuldnokk.** Militaarne retoorika. Argumentatsioon ja keeleline mõjutamine Eesti kaitsepoliitilises diskursuses. Tartu, 2011.

27. **Kai Tafenau.** Uue Testamendi tõlkimisest Rootsi ajal: käsikirjad, tõlkijad ja eesti kirjakeel. Tartu, 2011.
28. **Külli Prillop.** Optimaalsusteoreetiline käsitlus eesti keele fonoloogilisest kujunemisest. Tartu, 2011, 261 lk.
29. **Pärtel Lippus.** The acoustic features and perception of the Estonian quantity system. Tartu, 2011, 146 p.
30. **Lya Meister.** Eesti vokaali- ja kestkategooriad vene emakeelega keelejuhtide tajus ja häälduses. Eksperimentaalfoneetiline uurimus. Tartu, 2011, 145 lk.
31. **Kersti Lepajõe.** Kirjand kui tekstiliik. Riigieksamikirjandite tekstuaalsed, retoorilised ja diskursiivsed omadused. Tartu, 2011, 141 lk.
32. **Tiit Hennoste.** Grammatiliste vormide seoseid suhtlustegevustega eesti-keelses suulises vestluses. Tartu, 2013, 236 lk.
33. **Helena Metslang.** Grammatical relations in Estonian: subject, object and beyond. Tartu, 2013, 443 p.
34. **Kristel Uihoaed.** Verbiühendid eesti murretes. Tartu, 2013, 227 lk.
35. **Helen Plado.** Kausaalsuhete adverbiallaused eesti keeles. Tartu, 2013, 244 lk.
36. **Annika Kungas.** Pragmatiliste markerite kujunemine ja funktsioonid eesti keeles *lt*-sõnade näitel. Tartu, 2014, 200 lk.
37. **Maarika Teral.** Arvutipõhine eesti keele õpe: vahendid ja hinnangud nende efektiivsusele Tartu ülikooli keelekursuste näitel. Tartu, 2015, 175 lk.
38. **Anni Jürine.** The development of complex postpositions in Estonian: a case of grammaticalization via lexicalization. Tartu, 2016, 315 p.
39. **Kairi Tamuri.** Basic emotions in read Estonian speech: acoustic analysis and modelling. Tartu, 2017, 238 p.
40. **Sven-Erik Soosaar.** Eesti keele kultuursõnavara arengujooni – eesti keel kui piirikeel. Tartu, 2017, 157 lk.
41. **David Ogren.** Object case variation in Estonian *da*-infinitive constructions. Tartu, 2018, 212 p.
42. **Reet Hendrikson.** Kas sõjasõna sünnib sõtta? Erialakeele tõhusus sõjandusterminoloogia näitel. Tartu, 2018, 252 lk.
43. **Eva Velsker.** Ajaväljendid eesti murretes. Tartu, 2018, 216 lk.
44. **Riina Reinsalu.** Juhendavad haldustekstid žanriteoreetilises raamistikus. Tartu, 2019, 168 lk.
45. **Ann Siiman.** Vormikasutuse varieerumine ning selle põhjused osastava ja sisseütleva käände näitel. Tartu, 2019, 209 lk.
46. **Tiina Laansalu.** Kesk-Harjumaa kohanimed kujunemine ja nimesiire. Tartu, 2020, 216 lk.
47. **Mariko Faster.** Hargla kihelkonna vetenimed Eesti-Läti piirilal. Tartu, 2020, 218 lk.
48. **Maarja-Liisa Pilvik.** Action nouns in a constructional network: A corpus-based investigation of the productivity and functions of the deverbal suffix *-mine* in five different registers of Estonian. Tartu, 2021, 252 p.

49. **Hanna Pook.** Pronoun use and variation in Estonian dialects: *kes* 'who', *mis* 'what' and *keegi* 'someone'. Tartu, 2023, 225 p.
50. **Andra Annuka-Loik.** Naer ja naerev hääl eesti suulises suhtluses. Tartu, 2023, 202 lk.
51. **Tiina Paet.** Võõrainese kinnistumine eesti keeles: keelekorralduslik ja leksikograafiline vaade. Tartu, 2023, 260 lk.
52. **Lydia Risberg.** Sõnatähendused ja sõnaraamat. Kasutuspõhine sisend eesti keelekorraldusele. Tartu, 2024, 286 lk.
53. **Liis Ermus.** The phonetic variation of plosives in Estonian. Tartu, 2024, 151 p.
54. **Piret Upser.** Eesti hädaabikõned: tüübid, teemad ja suhtlusprobleemid. Tartu, 2025, 214 lk.