



Charlotte Stormbom

**Gendered language in flux:
The use of epicene pronouns in
EFL and ELF writing**



Charlotte Stormbom

Born 1989

Previous studies and degrees

Master of Arts, Åbo Akademi University, 2014

Bachelor of Arts, Åbo Akademi University, 2011

Picture credit: Max Askelo



Gendered language in flux: The use of epicene pronouns in EFL and ELF writing

Charlotte Stormbom

English Language and Literature
Faculty of Arts, Psychology and Theology
Åbo Akademi University
Åbo, Finland, 2021

Supervisors

Professor Tuija Virtanen-Ulfhielm
English Language and Literature
Åbo Akademi University

Professor Saara Haapamäki
Swedish Language
Åbo Akademi University

Reviewers

Professor Anna Mauranen
English Philology
University of Helsinki

Professor Anne Pauwels
Department of Linguistics
SOAS University of London

Opponent

Professor Anna Mauranen
English Philology
University of Helsinki

ISBN 978-952-12-4025-6 (printed)
ISBN 978-952-12-4026-3 (digital)
Painosalama Oy, Turku, Finland 2021

Acknowledgements

The completion of this dissertation would not have been possible without the help of numerous people. First and foremost, I wish to express my heartfelt gratitude to my supervisors Professor Tuija Virtanen-Ulfhielm and Professor Saara Haapamäki for your kind support and guidance along the way. I am also very grateful to my pre-examiners Professor Anna Mauranen and Professor Anne Pauwels.

I would like to thank everyone in the department of English Language and Literature: you have provided me with a positive and supportive research environment, even in the stressful time of a global pandemic.

I am very grateful to everyone who helped me distribute the elicitation experiment to university students across Europe: Professor Elena Borzova, Professor Jana Chamonikolasová, Dr Heidrun Dorgeloh, Professor Roberta Facchinetti, Professor Laure Gardelle, Dr Elisabeth Gustavsson, Professor Peter Harder, Dr Pekka Lintunen, Professor Satu Manninen, Dr Alessandra Molino, Professor Arja Piirainen-Marsh, Professor Natalia Pushina, Dr Janne Skaffari, Dr Erik Smitterberg, Professor Sandrine Sorlin, Dr Heli Tissari, Professor Jukka Tyrkkö, and Dr Jim Walker. Special thanks are due to Dr Iris Lindahl-Raittila for helping me pilot this experiment twice.

I want to thank Dr Tove Larsson for permission to use and collect data from the Swedish subset of the *Varieties of English for Specific Purposes dAtabase*. A sincere thank you is also owed to Dr Signe-Anita Lindgrén who encouraged me to start on this journey.

I gratefully acknowledge the financial support I have received while working on this project. This dissertation would not have been possible without funding from Stiftelsens vid Åbo Akademi H. W. Donners fond. The very final stage of the dissertation was completed with the help of a grant from the Åbo Akademi University Rector's Fund.

Finally, I wish to thank everyone who has enriched my life outside the academic world. Many thanks to my friends Annica, Linda, Lisa, Max, and Veijo for countless hours talking about anything and everything. Thank you to my family for believing in me: my mother Ildikó and my brothers Michael and Henry. Sadly, my father Christian did not live to see me finish this project, but I know he would have been the proudest of them all.

Thank you, Andrew, for your unending support and encouragement – and for your English expertise.

Åbo/Turku, 4 January 2021

Charlotte Stormbom

Abstract

The purpose of the dissertation is to investigate the use of epicene pronouns in written English as a foreign language (EFL) and English as a lingua franca (ELF). Epicene pronouns are defined as third-person singular pronouns used in reference to gender-indefinite antecedents. The dissertation focuses on the three most common types of epicene pronouns in English: generic *he*, *he or she* variants, and singular *they*.

The dissertation includes four substudies, which examine how the use of epicene pronouns has changed over time, how the antecedent affects the pronoun choice, and how the pronoun choice is influenced by the language users' L1 background. The materials consist of five corpora and an experiment. The corpus texts include: (1) argumentative essays written by university students from 13 L1 backgrounds, (2) academic papers written by L1 Swedish-speaking university students, and (3) articles published in Open Access (OA) journals.

The findings reveal diachronic changes in the use of epicene pronouns: the use of generic *he* is found to be considerably less common in the texts from the 2010s than in the texts from the 1970s to 2000s. At the same time, the use of singular *they* appears to be increasing in popularity in EFL and ELF writing.

Furthermore, the results show that the choice of epicene pronouns is affected by the antecedent type. Antecedents that are semantically plural, such as *everyone*, are more likely to co-occur with *they* than antecedents that are semantically singular, such as *a person*. The use of *he* and *he or she* displays the opposite kind of distribution. The gender expectancy of the antecedent also appears to have bearing on pronoun choice as generic *he* is more common with male-stereotyped nouns than with neutral or female-stereotyped ones.

The use of epicene pronouns is also influenced by the users' L1 lingua-cultural background. Speakers of L1s with pervasive grammatical gender (e.g. Italian and Russian) seem more prone to using generic *he* than speakers of genderless languages (e.g. Finnish) and languages with a combination of natural and grammatical gender (e.g. Swedish and Danish). A possible reason for these tendencies is that the structure of the L1 leads to differences in learners' conceptualisations of gender, in line with the framework of 'thinking for speaking', a weak form of linguistic relativity.

Keywords: epicene pronouns, singular *they*, corpus linguistics, linguistic relativity, gender, EFL, ELF

List of publications

This dissertation is a summary of the following four articles:

1. Stormbom, C. 2018. 'Epicene pronouns in intermediate to advanced EFL writing'. *International Journal of Learner Corpus Research* 4(1): 1-22.
<https://doi.org/10.1075/ijlcr.16016.sto>.
2. Stormbom, C. 2020. 'Singular *they* in English as a foreign language'. *Applied Linguistics Review*. Advance online publication.
<https://doi.org/10.1515/applirev-2019-0115>.
3. Stormbom, C. 2019. 'Language change in L2 academic writing: The case of epicene pronouns'. *Journal of English for Academic Purposes* 38: 95-105.
<https://doi.org/10.1016/j.jeap.2019.02.001>.
4. Stormbom, C. 2020. 'Gendering in open access research articles: The role of epicene pronouns'. *English for Specific Purposes* 60: 193-204.
<https://doi.org/10.1016/j.esp.2020.08.001>

Reprints of the articles were made with the kind permission of the publishers.

Table of contents

Acknowledgements.....	i
Abstract.....	iii
List of publications.....	iv
Table of contents.....	v
List of figures.....	viii
List of tables.....	viii
Abbreviations.....	ix
1 Introduction.....	1
1.1 Research background.....	3
1.2 Aims and research questions.....	4
1.3 Outline.....	5
2 Whose English?.....	6
2.1 The global spread of English.....	6
2.2 English as a foreign language (EFL) versus English as a lingua franca (ELF).....	7
3 The connection between gender, language, and thought.....	9
3.1 The expression of gender in language.....	9
3.1.1 Grammatical gender.....	9
3.1.2 Lexical gender.....	11
3.1.3 Social gender.....	11
3.2 Grouping languages according to the expression of gender.....	12
3.3 Gender bias in language.....	14
3.4 Gender and linguistic relativity.....	17
3.4.1 Overview.....	17
3.4.2 Previous research on grammatical gender and linguistic relativity..	19
3.4.3 Relating linguistic relativity and L2 use of language.....	20
4 Epicene pronouns in English.....	23
4.1 <i>He, he or she, or they?</i>	23
4.2 Epicene antecedents.....	26
4.3 Research on the use of and attitudes to epicene pronouns.....	28

5	Materials and methods.....	31
5.1	Corpus linguistics.....	32
5.1.1	Corpora.....	33
5.1.2	Extracting epicene pronouns.....	36
5.2	Experimental techniques.....	38
5.2.1	Experimental design	38
5.2.2	Data collection	40
5.3	Methods of analysis.....	40
6	Summaries of the studies.....	43
6.1	Article 1: Epicene pronouns in intermediate to advanced EFL writing....	43
6.2	Article 2: Singular <i>they</i> in English as a foreign language.....	44
6.3	Article 3: Language change in L2 academic writing: The case of epicene pronouns.....	45
6.4	Article 4: Gendering in open access research articles: The role of epicene pronouns.....	45
7	Results.....	47
7.1	Pronoun types across data sets.....	47
7.2	Variation according to antecedent.....	48
7.2.1	Antecedent type	48
7.2.2	Gender stereotyping.....	51
7.3	Variation according to research material	52
7.3.1	Student academic writing versus elicited data	52
7.3.2	Student academic writing versus professional academic writing	53
7.4	Variation according to L1 lingua-cultural background	54
7.4.1	Overall findings	54
7.4.2	Antecedent type and gender stereotyping.....	57
7.5	Variation at the individual level.....	57
7.6	Summary of the results	59
8	Discussion.....	62
8.1	The use of epicene pronouns in EFL and ELF writing: A synthesis	62
8.2	Implications.....	66
8.3	Limitations.....	70

9 Conclusion.....	72
Svensk sammanfattning.....	73
References	75
Appendix: Original publications	85

List of figures

Figure 1. Spectrum of gender expectancy in nouns.....	60
Figure 2. Factors affecting the outcome of feminist language reforms. Adapted from Milles (2013: 115).....	65

List of tables

Table 1. Formal expression of gender in languages.....	13
Table 2. Types of epicene antecedents and their notional number	26
Table 3. Aspects of individuation	27
Table 4. Details of the four substudies.....	31
Table 5. Corpora used in the dissertation.....	34
Table 6. Search words and context words in WordSmith Tools.....	37
Table 7. Antecedents in the experiment in Study 2	39
Table 8. Pronoun types in the overall data sets.....	48
Table 9. Distribution of pronouns according to type of antecedent in the corpus materials.....	49
Table 10. Distribution of pronouns according to antecedent in the experiment.....	50
Table 11. Comparison between L1 speakers of Swedish in the experiment and in BATMAT/VESPA-SE.....	52
Table 12. Comparison between student academic writers and professional academic writers.....	53
Table 13. Comparison between L1 groups in the ICLE and in the experiment..	55
Table 14. Modified paradigm of personal pronouns in English.....	70

Abbreviations

COAJA	Corpus of Open Access Journal Articles
EAP	English for Academic Purposes
EFL	English as a Foreign Language
ELF	English as a Lingua Franca
ICLE	International Corpus of Learner English
L1	First language
L2	Additional (foreign) language
LOCNESS	Louvain Corpus of Native English Essays
VESPA	Varieties of English for Specific Purposes dAtabase

1 Introduction

Every year since 2003, the American dictionary company Merriam-Webster has published a list of ten words of the year from the English language. In 2019, a personal pronoun made it to the very top of this list: singular *they*. The use of this pronoun is two-fold: on the one hand, it can be used in co-reference with singular antecedents of indeterminate or irrelevant gender; on the other hand, it may reference individuals who do not identify as either male or female (see e.g. Baron 2020; Bradley et al. 2019). This dissertation is primarily concerned with the first of these functions, in which singular *they* can be categorised as an ‘epicene pronoun’ (Baron 1981; 1986). Example (1) illustrates the use of singular *they* as an epicene pronoun (1a), as well as the use of two other common types of epicenes in English: generic *he* (1b), and coordinations and composites like *he or she* and *he/she* (1c). The co-referential antecedent in this example is the gender-indefinite noun phrase (NP) *a person*.

- (1) a. Coercion means forcing *a person* to do something against *their* will.
- b. Coercion means forcing *a person* to do something against *his* will.
- c. Coercion means forcing *a person* to do something against *his/her* will.

Like most languages, English does not have a specific singular pronoun whose *only* function would be to reference a generic individual, that is, an individual of unknown or indeterminate gender (see Corbett 1991). Instead, the tradition in English has been to use personal pronouns that also perform other functions. The types illustrated in (1), singular *they*, generic *he*, and *he or she* variants, have been shown to be the most frequent epicene pronouns in present-day native (L1) English (see e.g. Newman 1998; Laitinen 2007; Paterson 2014), as well as in English as a foreign language (EFL) writing according to preliminary research conducted for the present dissertation. The construction in (1) can be described as a case of anaphora, where a singular gender-indefinite NP or a singular indefinite pronoun co-occurs with an epicene pronoun. Corresponding cases of cataphora, as exemplified in (2), are not considered further here as cataphora is a relatively rare phenomenon in English (see e.g. Halliday & Hasan 1976; for a recent study of cataphora, see Laitinen 2011).

- (2) Without a clear picture of what is expected of *him*, *a child* can become frustrated and overwhelmed.

Traditional grammar describes pronouns as ‘placeholders for nouns’ that do not contribute with any meaning of their own. Yet, this understanding of pronouns is questionable: for instance, studies have shown time and again that the pronoun *he* cannot be used to denote a generic individual because it evokes male images in the language user’s mind (for a review, see Henley & Abueg 2003). Moreover, the successful introduction of a new gender-neutral pronoun in Swedish – *hen* – goes to show that pronouns form a more flexible category

than previously assumed. In recent times, epicene pronouns have featured frequently in discussions about gender and gender-fair language in English; these pronouns raise important questions about who is allowed representation in language and as such their use is an important matter to many people. The past few decades have also seen plenty of scholarly interest in the use of epicene pronouns in English. Nonetheless, the vast majority of previous studies have focused on L1 varieties. While studies of L1 use are valuable in their own right, it cannot be ignored that L1 speakers of English are now outnumbered by speakers for whom English is an additional (L2) language¹ (see e.g. Crystal 2003). The small number of existing L2 studies show that L2 speakers do not necessarily follow the same patterns as L1 speakers in their use of epicene pronouns, particularly not as regards the frequency of generic *he* versus the frequency of singular *they* (Abudaljuh 2012; Lee 2007). Thus, the present dissertation extends previous research by mapping how L2 speakers from a variety of L1 backgrounds use epicene pronouns in English writing. The dissertation also examines how epicene pronouns are used when English serves as a global means of communication, in written English as a Lingua Franca (ELF).

Previous research has shown that the use of epicene pronouns in L1 English may be influenced by a variety of factors, such as the level of formality of the context (see e.g. Curzan 2014; Hekanaho 2020; Holmes 1998; Parini 2012) as well as the type of antecedent with which the pronoun co-occurs (see e.g. Baranowski 2002; Gerner 2000; Paterson 2014, 2020). In the context of L2 use, another potentially influential factor is the L1 lingua-cultural background². The languages of the world display immense variation in terms of how they express gender. In some languages, gender is marked on every single noun as well as on all elements agreeing with the nouns, while other languages have no formal marking of gender at all, not even on pronouns like English does. Given this situation, it can be expected that L2 users of English will approach the issue of epicene pronouns in different ways depending on their L1 lingua-cultural background. One point of departure in understanding influence from the L1 is Slobin's (1996) framework of *thinking for speaking*, a weak form of linguistic relativity which holds that structural differences in language lead to differences in language-mediated cognition. For example, speakers of L1s with pervasive grammatical gender are likely to be more aware of and attentive to gender than L1 speakers of genderless languages (i.e. languages that lack grammatical gender) because their L1 requires them to constantly access gendered forms.

¹ The abbreviation L2 is used throughout the dissertation to denote an additional language, regardless if it is the language user's second or nth language in order of acquisition.

² Following Friedrich (1989: 306), *lingua-culture* is here understood as "a domain of experience that fuses and intermingles the vocabulary, many semantic aspects of grammar, and the verbal aspects of culture".

1.1 Research background

The study of language and gender has seen different trends over the past sixty years. The beginnings of this field of research were strongly tied to the second wave of feminism, which had its roots in the United States in the late 1960s and early 1970s. The movement quickly reached other parts of the Western world, including the United Kingdom, France and Germany. Second-wave feminist linguists were particularly interested in how structural aspects of language result in gender bias or sexism, primarily by excluding women as referents. Many studies within this tradition examined how language users interpret masculine generics, such as generic *he* in English (see e.g. Hamilton 1988; MacKay & Fulkerson 1979; Martyna 1978; Moulton et al. 1978). These studies showed that masculine generics are not gender-inclusive, as they tend to be interpreted as referring to men only (for a review, see Henley & Abueg 2003; see also Bußmann & Hellinger 2001, 2002, 2003; Hellinger & Motschenbacher 2015). The linguistic work tied to second-wave feminism also examined stereotypical speech of men and women, primarily focusing on gender differences and how such differences can lead to miscommunication between the genders (see e.g. Lakoff 1975; Spender 1980; Tannen 1991).

A very central change that third-wave feminism brought with it was a new way of conceptualising gender (see e.g. Mills 2003). Moving away from an essentialist view, third-wave feminism regards gender as a performative construct, something we as humans do in interaction (see Butler 1990). From this follows that third-wave feminist linguists do not view women as a homogeneous group, but instead emphasise the fact that women's speech is highly diverse (see e.g. Coates & Cameron 1988; Eckert 2000). Third-wave feminist linguistics primarily focuses on gender and sexism as discursive practices that are constructed and performed locally, rather than on the global concerns of second-wave feminism.

In English language communities, the study of language and gender has shifted its focus from second-wave structurally oriented analyses of sexism to third-wave discursive investigations of gendered language. This does not mean, however, that there is no longer a place for research that examines structural aspects of language. Pauwels (2011: 11) notes that "in speech communities that have only recently started looking into gender-based linguistic discrimination, the main focus of attention is on the lexical and grammatical elements marking sexism". This is worth bearing in mind in the context of English as a global language as English language users come from so many L1 lingua-cultural backgrounds with different linguistic features and approaches to gender. Following for instance Mills (2003), Motschenbacher (2016), and Pauwels (2011), the stance taken in this dissertation is that the study of gender and language in EFL and ELF is best pursued by combining structural and discursive approaches. Thus, approaches from both second and third-wave feminist linguistic traditions are incorporated in the four substudies.

1.2 Aims and research questions

The goal of this dissertation is to investigate how epicene pronouns are used in EFL and ELF writing. The following questions serve as a basis for the dissertation as a whole:

1. What differences, if any, can be seen between groups of learners with different L1s in terms of the use of epicene pronouns in English writing? (Study 1 and Study 2)
2. How, if at all, do different aspects of the antecedent (e.g. definiteness, notional number, gender-stereotyping) affect the choice of epicene pronouns in EFL and ELF writing? (Study 1, Study 2, Study 3, and Study 4)
3. What kinds of, if any, diachronic changes can be observed in the use of epicene pronouns in L2 English since the 1970s? (Study 3)

In addition to these overall aims, each of the four articles has its own specific research questions, which are addressed in the summary of the studies in Chapter 6. In the studies focusing on EFL (Studies 1-3), all learners are students who major in English Language and Literature at universities across Europe. While these students represent a rather specific group of English learners, their use of epicene pronouns is relevant to consider for two main reasons. Firstly, it can be assumed that individuals studying to become language professionals are sensitive to linguistic phenomena in general. It is therefore of particular interest if gender-biased forms are used persistently in some of the L1 subgroups. For instance, evidence of influence from the L1 would suggest that gender is a highly salient category even in individuals who are immersed in the L2 and have a strong metalinguistic awareness, evoking weak models of linguistic relativity like Slobin's (1996) thinking for speaking. Another incentive for examining students of English is that many of these individuals will go on to be teachers of English at various levels – and they will then serve as models for future learners of the language. Thus, it is both noteworthy and concerning if gender-biased forms are frequently reproduced by this group of students.

The study of ELF writing (Study 4) examines research articles published in two academic fields: (1) languages and linguistics, and (2) library and information science. One reason for focusing on articles in these fields is that they frequently include references to generic individuals; in many other academic disciplines, this is not the case because of the nature of the studies. Moreover, the set of referents tend to be similar in the two disciplines (e.g. *user*, *student*), which makes comparisons between them feasible. Nonetheless, it can be assumed that language professionals are generally more sensitive to linguistic matters, which could potentially lead to differences in how the issue of epicene pronouns in English is addressed.

1.3 Outline

The dissertation is structured in the following way: Chapter 2 focuses on English in a global world, examining how the concepts of English as a foreign language (EFL) and English as a lingua franca (ELF) relate to each other. In Chapter 3, the focus shifts to the expression of gender in language. The chapter explores how gender is linguistically realised and how the expression of gender may lead to bias in language. It also includes a discussion of the influence of gender on language-mediated cognition. Chapter 4 provides an in-depth examination of the issue of epicene pronouns in English. A description is given of the different types of epicene pronouns and the types of antecedents with which they co-occur. The chapter ends with an overview of previous studies. Chapter 5 presents the materials as well as the methods of data collection and analysis. Following this, in Chapter 6, the four articles that the dissertation is based on are summarised. Chapter 7 provides the results of the studies and relates them to the findings of previous research. In Chapter 8, a synthesis is given of the use of epicene pronouns in EFL and ELF writing. The chapter also discusses the implications and limitations of the research and provides suggestions for future research. The final chapter of the dissertation, Chapter 9, offers some concluding remarks on the work.

2 Whose English?

This chapter explores the global spread of English and its implications for how the language is conceptualised in research. Two central concepts are discussed in depth: (1) English as a foreign language and (2) English as a lingua franca.

2.1 The global spread of English

The English language occupies a unique position in today's world, having developed into a truly global lingua franca; the number of L2 speakers of English now exceeds that of L1 speakers of the language (see e.g. Crystal 2003). A commonly used description of the diffusion of English around the world is Kachru's (1985) framework of three concentric circles: the inner circle, the outer circle, and the expanding circle. The regions in these circles differ in the patterns of acquiring the language, as well as what functional domains the language holds. The inner circle refers to regions where English has traditionally been the native language, such as the United Kingdom, the United States, Australia, and New Zealand. The outer circle, in turn, includes regions where the English language has official status in some form or other, such as India, the Philippines, Nigeria, and Kenya. Finally, the expanding circle is composed of regions in which English is not institutionalized, but in which the language is still widely used as a foreign language and as a lingua franca. This includes a great part of the world's population, for instance China, Russia, and most of the European countries.

The global nature of English raises questions about norms and standardization – ownership of the language, in short. The relationship between the three concentric circles in Kachru's model has traditionally been hierarchical, with the inner circle varieties being the providers of norms of language use. Nonetheless, norm orientation may shift over time; already in 1985, Kachru (1985: 30) noted that “the global diffusion of English has taken an interesting turn: the native speakers of this language seem to have lost the exclusive prerogative to control its standardization”. Since then, the Internet and social media have further strengthened the position of English as a global language (see Lee 2017). Thus, while Kachru's model continues to be a standard point of reference, many additional models have since been proposed to account for the complexity in the use of English as a global language (see e.g. Mair 2013; Modiano 1999; Schneider 2007, 2014). For example, Modiano (1999: 26) suggests a model with three circles in which the inner circle consists of speakers who are “proficient in international English”. In other words, these language users are not necessarily L1 speakers of English, but rather speakers of any varieties that are conducive to cross-cultural communication. The next level in the model includes speakers whose use of English functions successfully in local contexts but not in cross-cultural communication, again including both native and foreign language proficiency. The outer circle, lastly, consists of EFL learners whose knowledge of the languages is highly limited. This type of model challenges the traditional view of English, according to which L1 speakers are the only providers of norms.

Nonetheless, the degree to which L1 standards are perceived as the ultimate target will vary depending on the context in which English is used. With respect to the present work, it should be kept in mind that both EFL teaching and academic publishing tend to be conservative and very oriented towards L1-speaker norms (see e.g. Turner 2018).

2.2 English as a foreign language (EFL) versus English as a lingua franca (ELF)

The present dissertation is primarily concerned with English in Kachru's expanding circle. Given the vast spread of the language, this circle is very heterogeneous, representing many varieties of English. Moreover, there is enormous variation in terms of what status the English language has in the different regions in the expanding circle. English has had a historically high profile in many parts of northern Europe, for instance, in Germany, the Netherlands, and the Scandinavian countries, whereas this is not the case in some other European countries, such as Spain and Slovakia (see Lillis and Curry 2006: 6).

A distinction is often made between learners of English as a foreign language (EFL) on the one hand, and users of English as a lingua franca (ELF) on the other hand (for a discussion of these terms, see Mauranen 2018). A central baseline for this distinction is the context in which English is used. EFL is strongly associated with a classroom setting of some form, where the EFL learners usually share the same L1 background. Moreover, learners and teachers of EFL are generally guided by norms set by L1 varieties, that is, norms of what tends to be referred to as 'Standard English'. The EFL learner is thus a representative of the expanding circle of English in Kachru's model. As far as research is concerned, studies of EFL typically control for language proficiency and L1 background of the language learners, and comparisons between L1 groups as well as between L2 and L1 speakers are commonplace (see e.g. Granger 2015).

Although useful in learner contexts, the concept of EFL appears insufficient in describing international use of English, which includes all three concentric circles in Kachru's model of English. Thus, in recent years, researchers have given increasing attention to the use of English as a lingua franca (ELF). A well-known, early definition of ELF is Firth's (1996: 240) description of ELF as "a 'contact language' between persons who share neither a common native tongue nor a common (national) culture, and for whom English is the chosen foreign language of communication". A distinctive feature of ELF is therefore that the participants in ELF contexts are fluent in English to varying degrees. Based on Firth's definition, L1 speakers can also be included in ELF situations since they do not share a common language with L2 speakers of English. However, not all researchers agree that L1 speakers should be included in ELF (see e.g. House 1999: 74; Phillipson 2008: 263), nor is there any consensus as to what exactly ELF is (see e.g. Prodromou 2007). Some have even argued that ELF is a language

in its own right, or, at the very least, a new variety of English (see e.g. Jenkins 2006; Leung 2005; Seidlhofer 2004). Others, again, claim that ELF is best conceptualised as a “dynamic context of use” (Park & Wee 2012: 59). The present dissertation follows the latter definition, that is, ELF is viewed as an activity type where the goal is to communicate successfully in a context where there is no shared L1 between the participants.

Both EFL use and ELF use are included in the substudies of the dissertation: The first three articles examine the use of epicene pronouns in what can be labelled as EFL. The participants in Studies 1-3 are EFL learners in the specific contexts in which the data were collected, that is, in departments of English at universities across Europe. However, in their private lives and in their (future) professional lives, these students might equally well be called users of ELF. In other words, the distinction between EFL learner and ELF user is not always clear-cut and both labels may be attached to one and the same individual depending on the context. As Swan (2012: 388) concludes: “EFL teachers have students many of whom will use what they have learnt for *lingua franca* purposes. In this sense, EFL leads to ELF; they are on opposite sides of the same coin.” In the fourth study, the focus shifts to English in a global context: professional academic writing. The use of English in this type of context can best be described as ELF, as English functions as a common means of communication between individuals from a wide range of cultural and linguistic backgrounds. A more specific label for this kind of use of English is English as an academic *lingua franca* (see e.g. Ingvarsdóttir & Arnbjörnsdóttir 2013; Mauranen 2012; Mauranen, Hynninen & Ranta 2010; Tardy 2004). Again, the categorisation of academic writers as ELF users is not an absolute one in the sense that these same individuals may be learners of English in some other context of their lives. It should also be noted that publishing in English has traditionally been, and still is, very focused on L1 standards, with academic journals strongly recommending L1 proofreading of manuscripts (see e.g. Turner 2018).

3 The connection between gender, language, and thought

The purpose of this chapter is to discuss the concept of gender and how it may be constructed through language. The chapter also explores whether, and if so how, formal expression of gender in language can influence how we think about the world.

3.1 The expression of gender in language

3.1.1 Grammatical gender

Gender can be conveyed in various ways in language. A highly pervasive form of gender expression is ‘grammatical gender’, which refers to a noun classificatory system that affects agreement to various degrees. Depending on the language, grammatical gender steers agreement with different satellite elements, such as adjectives, articles, anaphoric pronouns, and verbs. For instance, (3) shows an example from French (from Schafroth 2003: 90), where the feminine noun *voiture* ‘car’ affects the form of a number of elements: the possessive pronoun *sa* ‘his’ (masculine form *son*), the adjective *nouvelle* ‘new’ (masculine form *nouveau*), and the direct object pronoun *la* ‘it’ (masculine form *le*).

- (3) *Pierre adore sa nouvelle voiture*
Pierre adores POSS.3.SG.FEM new.SG.FEM car.SG.FEM
et il la montre à tout le monde.
and he 3.SG.FEM.ACC shows to everyone
‘Pierre adores his new car and shows it to everyone.’

While gender assignment may seem highly arbitrary at a first glance, there are generally various patterns of how nouns are categorised into gender. For instance, nouns with a similar morphology may be grouped together. An example of this can be seen in German, where all nouns with the diminutive suffixes *-chen* and *-lein* belong to the neuter gender. In addition, there is always a semantic core to gender assignment, which in many languages is biological gender or sex. Thus, the basis in Indo-European languages tends to be that nouns that refer to female individuals, i.e. lexically female nouns, are grammatically feminine, while nouns that are lexically male are grammatically masculine. To give an example, the German noun *Mann* ‘man’ is assigned to the masculine gender, whereas the noun *Frau* ‘woman’ is feminine. This difference in gender is noticeable, for instance, in the fact that the two nouns take different definite articles, *der* for masculine and *die* for feminine. Grammatical gender classifications can have both a formal function and a semantic function, in cases where the referent’s sex is denoted.

Although there is usually a correspondence between grammatical gender and biological gender, there can also be exceptions to this general rule. For instance,

the German noun *das Mädchen* ‘the girl (n)’ belongs to the grammatical gender neuter, although one would expect it to be feminine based on the biological gender of the referent. Some nouns of this kind are so-called ‘hybrid nouns’, which means that semantic gender may override grammatical gender in agreement in certain cases. An example from Corbett (1991: 228) is given in (4).

- (4) *Shau dir dieses Mädchen an, wie gut*
look you this.NEUT girl at how good
sie/es Tennis spielt
she/it tennis plays
‘Look at this girl, how well she plays tennis.’

Based purely on grammatical gender, the neuter personal pronoun *es* would be expected in this sentence, but it is also common to use the feminine pronoun *sie*, which agrees with the noun in terms of semantic gender. The choice between grammatical and semantic gender can partly be predicted by the agreement hierarchy proposed by Corbett (1978: 204). The syntactic positions of the hierarchy are arranged in the following way:

attributive – predicate – relative pronoun – personal pronoun

If syntactic agreement occurs with one of these controllers, it will also do so with all controllers to the left of it in the hierarchy. Conversely, if semantic agreement is possible with one controller, it will be possible with all elements to the right of it. The further to the right of the hierarchy one moves, the more likely it is that semantic rather than syntactic agreement will occur. Thus, it is more likely that a personal pronoun, as in example (4) above, will take semantic agreement, than that a relative pronoun will do so.

In many gendered languages, there are asymmetries in the use of the feminine and the masculine forms of personal nouns. Whereas the feminine form can only refer to female individuals, the masculine form typically has three potential uses: (1) in reference to men, (2) in reference to a group consisting of both men and women, and (3) in generic reference, that is, when the gender of the referent is indefinite or unknown (see e.g. Bußmann & Hellinger 2001, 2002, 2003). For instance, the feminine plural noun *cittadine* ‘citizens (f)’ in Italian can only be used in reference to female citizens, whereas the masculine form, *cittadini* ‘citizens (m)’, may perform all three functions mentioned above (see e.g. Marcato & Thüne 2002).

Research has shown that the use of masculine generics in languages with grammatical gender is biased as masculine generics tend to be associated exclusively with men in individuals’ mental representations (for a review, see Gygax et al. 2009; see also Braun et al. 2005 on German; Gygax et al. 2012 on French; Schmid 1998 on Russian). This effect seems to occur even when language users have explicitly been taught the rule that the masculine form includes individuals of both sexes (see Gygax et al. 2009: 242).

3.1.2 Lexical gender

A lack of structural linguistic markers of gender does not mean that gender is not or cannot be expressed in a given language; as Jakobson (1959: 236) so aptly puts it, “[l]anguages differ essentially in what they *must* convey and not in what they *may* convey”. A universal semantic feature of language is lexical gender, which refers to linguistic forms that carry the denotation [male] or [female] (Corbett 1991: x). In English, common nouns of this kind include kinship terms, such as *brother* and *mother*, as well as some other basic nouns denoting human beings, such as *boy* and *woman*. Lexical gender is sometimes also formally marked, as in the English noun *actress* (female suffix *-ess*), or the Swedish noun *lärarinna* ‘female teacher’ (female suffix *-inna*). As mentioned, lexical gender tends to converge with grammatical gender in languages with pervasive gender. For example, the German nouns *Junge* ‘boy (m)’ and *Mann* ‘man (m)’ are lexically male as well as being grammatically masculine. Languages that lack formal marking of gender still have lexical gender, although this may be less obvious than in languages that have some form of gender agreement, such as the use of the gendered personal pronouns *he* and *she* in English. In the genderless language Finnish, an example of a lexically gendered word is the noun *kuningas* ‘king’, which is clearly lexically male in that it can only reference a male individual.

3.1.3 Social gender

Social gender refers to the roles and norms ascribed to men and women in society. Like lexical gender, social gender as a semantic category is found universally in language as all languages have nouns that are stereotypically associated with maleness or femaleness; in Aikhenvald’s (2016: 74) words, stereotypes of social gender “will be reflected in any language – with or without Linguistic Gender – as long as they are relevant for the community which speaks it”. In English, two examples of stereotyped nouns are *nurse* (high female gender-expectancy) and *pilot* (high male gender-expectancy) (see e.g. Kennison & Trofe 2003; Oakhill et al. 2005). Stereotypes may be specific to a given language and culture, but many stereotypes are ubiquitous in Western societies. For instance, Misersky et al. (2014) examined stereotyping of nouns in six languages spoken across Europe and in Northern America: Czech, English, French, German, Italian, Norwegian, and Slovak. Nouns that proved to have a high gender-expectancy in all six languages included the male stereotyped nouns *president* and *drug dealer*, and the female-stereotyped nouns *social worker* and *flight attendant*. Neutral personal nouns were, for instance, *kid*, *student*, *dog owner*, and *environmentalist*.

Stereotyping in nouns is included in what McConnell-Ginet (2008: 512) refers to as ‘conceptual baggage’:

I include under this rubric what traditional lexicographers and others have called connotations, but also encyclopedic knowledge, stereotypes or prototypes, and background assumptions, as well as knowledge about social

practices in the course of which the word gets used. [...] I treat conceptual baggage as part of lexical significance because it can have profound communicative effects, triggering various kinds of (virtually automatic) inferences by interpreters, inferences not always explicitly recognized by interpreters as such and often not intended by speakers.

In a similar fashion, generic masculines (see above) carry conceptual baggage in the sense that they automatically trigger the meaning 'male individual', rather than 'individual of any gender'.

3.2 Grouping languages according to the expression of gender

Following the taxonomy presented in Gyga et al. (2019), languages can be divided into five main groups depending on the formal expression of gender: (1) genderless languages (e.g. Finnish, Turkish), (2) genderless languages with traces of grammatical gender (e.g. Basque), (3) languages with natural gender (e.g. English), (4) languages with a combination of natural and grammatical gender (e.g. Swedish, Danish), and (5) languages with grammatical gender (e.g. French, German, Italian) (see also Corbett 1991). Languages belonging to category (2) will not be discussed further in this dissertation as they do not occur as L1s in any of the substudies. Table 1 shows an overview of the relevant groups of languages, including features that they display in relation to gender.

Table 1. Formal expression of gender in languages

Gender system	Traits	L1s represented in the dissertation
Genderless	No grammatical markers of gender	Finnish, Turkish
Natural gender	Anaphoric gender in pronouns, but no gender agreement on other elements	English
Natural and grammatical gender	Nouns are categorised into grammatical gender, but there are separate pronominal forms for grammatical and natural gender	Danish, Dutch*, Norwegian*, Swedish
Grammatical gender	All nouns are categorised into grammatical gender, and gender affects agreement with a variety of elements	Bulgarian, Czech, Dutch*, French, German, Italian, Norwegian*, Polish, Russian, Spanish

* In some varieties of Dutch and Norwegian, masculine and feminine gender have merged into a common gender class as in Danish and Swedish, whereas other varieties have retained the three-gender system with masculine, feminine, and neuter.

In genderless languages, there is no grammatical marking of gender, either on nouns or on any satellite elements. Thus, the Finnish language makes no distinction between male and female in pronouns – the third-person singular pronoun *hän* ‘he/she’ is used regardless of the sex of the referent; in Turkish, another genderless language, the corresponding third-person singular pronoun is *o* ‘he/she’ (see e.g. Lewis 2000). As described above, a lack of formal gender does not mean that genderless languages do not express gender at all; like all languages, genderless languages have lexical gender and social gender. For instance, the Finnish noun *sairaanhoitaja* ‘nurse’ is socially gendered in that language users tend to associate this noun with female referents (see Engelberg 2016).

Languages with natural gender have anaphoric gender in personal pronouns, but no other grammatical elements that show gender agreement. A language of this kind is English, which makes a distinction between the third-person singular pronouns *he* and *she*. This distinction is, with a few exceptions, solely based on natural gender, or sex. Most personal nouns in English are generic, which means that they are neither formally nor lexically gendered. Examples of such generic

nouns include *speaker* and *child*. When personal nouns are used referentially, the choice of co-referential pronouns is thus based on the (presumed) sex of the referent.

Languages with natural and grammatical gender fall somewhere in between languages like English and languages with pervasive gender, such as German. The Scandinavian languages Danish, Norwegian, and Swedish are included in this group with mixed gender, as is the Dutch language³. In some varieties of these languages, a three-gender system with masculine, feminine, and neuter gender still exists, and such varieties are consequently categorised as languages with pervasive grammatical gender, as described below. However, in other varieties, such as standard Swedish and standard Danish, there are only two genders: common and neuter. In this kind of gender system, the choice of third-person pronouns follows both natural gender and grammatical gender (see e.g. Corbett 1991: 247). In effect, this means that one set of pronouns is used in personal reference and another one in non-personal reference (see e.g. Andersson 2000). In Swedish and Danish, for instance, the pronouns *han* 'he' and *hon* (Sw.) / *hun* (Da.) 'she' are used almost exclusively in reference to humans (and sometimes in reference to animals), whereas the pronouns *den* and *det* 'it' are used with non-human antecedents. This situation differs from languages with grammatical gender only as such languages use the same sets of pronouns regardless if the noun in question is animate or inanimate, or human or non-human.

In languages with pervasive grammatical gender, lastly, all nouns are categorised according to gender. The number of genders varies, but in Indo-European languages they are typically two (masculine and feminine), or three (masculine, feminine, and neuter). Gender is always marked on several elements, which, again, may vary somewhat between languages. For instance, both Russian and German mark gender on adjectives, determiners, and pronouns, but, unlike German, Russian also has gender marking on some verb forms (see e.g. Bußmann & Hellinger 2001, 2002, 2003; Hellinger & Motschenbacher 2015).

3.3 Gender bias in language

The formal expression of gender may lead to various kinds of gender-bias in language use. Analyses of gender-bias have been undertaken in many linguistic communities, focusing on various language phenomena in different text types, including educational materials, grammars and dictionaries, mass media texts, and legislative and legal documents (for an overview, see Pauwels 1998). As shown in these studies, the tendency in many languages has been for women to be disfavoured by the conventions of language, reflecting a history of male dominance in society.

³ In some categorisations, Danish, Dutch, Norwegian, and Swedish are grouped as languages with natural gender, together with English (see e.g. Stahlberg et al. 2007).

A central issue that second-wave feminist linguistics focused on was the use of male forms not only in reference to male individuals, but also in reference to generic classes of human beings. Such practices include the use of generic *he* in English and generic *han* 'he' in Danish and Swedish, as well as the use of generic masculines in languages with grammatical gender (see also 3.1.1.). Example (5) illustrates the use of generic masculines in German (Bußmann & Hellinger 2003: 158).

(5) *Jeder (m) Wähler (m) sollte von seinem (m) Stimmrecht Gebrauch machen.*
'Every (m) voter (m) should exercise his right to vote.'

In this example, the head noun *Wähler* 'voter' is masculine, which steers the agreement of two other elements: *jeder* 'every' and *seinem* 'his' – both in the masculine form. The use of generic masculines can result in a pervasive male bias in gendered languages as all satellite elements agreeing with the head noun also need to be in the masculine form, as seen in (5). Studies have confirmed that generic masculines tend to be interpreted as male-specific (for a review, see Gygax et al. 2009).

Another form of gender bias in language is the existence of lexical asymmetries. It is common for languages to have corresponding male and female word pairs that differ diametrically in their connotations. In many cases, the female word has acquired negative meanings that extend well beyond what Baron (1986: 115) terms "the simple recording of gender". Two examples in English are the word pairs *master/mistress* and *bachelor/spinster*, where the male and female terms have very different associations. In languages with grammatical gender, the male form of a role name might even be preferred over the feminine one in reference to women, simply because the feminine form has lower status. For instance, a female Russian author will refer to herself using the masculine generic *avtor* 'author (m)' rather than the feminine form *avtorsha* 'author (f)' as the feminine form is derogatory (Aikhenvald 2016: 193). Similar examples can be found in other gendered languages as well: in Italian, many traditional generic masculines for professions, such as *avvocato* 'lawyer (m)', enjoy a much higher status than the feminine variants, such as *avvocatessa* 'lawyer (f)', and therefore tend to be preferred even when the referent is explicitly female (see Merkel et al. 2012).

The realisation that language can perpetuate gender bias has led feminist linguists across language communities to propose changes in the use of language. Such reform efforts are crucially based on the idea of a bidirectional relationship between social change and language change. Hellinger (1991: 25) describes this in the following way:

Nonsexist language is seen as an instrument of social change: not only does it reflect tendencies towards a more equal treatment of the sexes, it also provides speakers with alternative models (concepts) of reality, thus contributing to changes in attitudes and social practices.

This kind of thinking is in many ways in line with linguistic relativity, that is, the idea that language can influence thought in various ways (see 3.4). Feminist linguistic reform has typically taken the form of ‘corpus planning’ (see e.g. Pauwels 1998), which can be described as prescriptive measures taken to modify the language in various ways. Such modifications involve, for instance, the introduction of new lexical units as well as suggestions for changes in grammatical practices (see Kloss 1969). Two main strategies tend to be promoted for avoiding gender-bias in language use: ‘feminisation’ and ‘neutralisation’. While feminisation seeks to increase female visibility in language use, the goal of neutralisation is to avoid specification of gender where possible. An example of feminisation is the use of coordinated NPs in German, such as *Studenten und Studentinnen* ‘male students and female students’, instead of using the masculine form, *Studenten*, as a generic term. In turn, an example of neutralisation in English is to opt for the gender-indefinite noun *police officer*, rather than using the gendered nouns *policeman* and *policewoman*. Although both strategies may be found in one single language, there is a strong link between the linguistic make-up of a language and the type of strategy that is primarily employed. In languages with a pervasive system of grammatical gender, neutralisation is often not possible to achieve, and feminisation therefore tends to be the preferred solution. Nonetheless, an issue with feminisation strategies is that they can result in very cumbersome expressions if many elements show gender agreement with the noun. For instance, consider the Czech sentence in (6), where splitting of feminine and masculine forms is used (Čmejková 2003: 49).

(6) *Čtenář/ka (m/f) je vyzýván/a (m/f), aby sám/sama (m/f) odhalil/a (m/f) důsledky revidování jazykového paradigmatu pro text.*

‘The reader is invited to reveal for himself/herself the consequences of revising language paradigms for the text.’

Another issue with feminisation strategies is that they inexorably sustain the binary division of gender (see e.g. Baron 2020).

In languages where gender is structurally less salient, neutralisation is often preferred over feminisation. For instance, in English and Swedish, there has been a trend to avoid unnecessary specification of gender in nouns, with the result that female suffixes like *-ess* in English and *-inna* in Swedish have mostly fallen out of favour. Another aspect of language that can be neutralised is the use of personal pronouns. The Swedish language presents a particularly interesting case here as a new gender-neutral pronoun, *hen*, was introduced into the language in the 2010s. Pronouns are usually classified as closed-class items, which means that new members are rarely added to these classes – in contrast to open-class items, like nouns, which readily accept additions of new words. Despite this situation, the use of the pronoun *hen* spread quickly in Swedish

language use, and in 2015 the pronoun was added to *Svenska Akademiens ordlista* (SAOL 2015), the glossary of the Swedish Academy.

3.4 Gender and linguistic relativity

At the heart of feminist language reform lies the notion that language can affect how we perceive things; thus, from this standpoint, an unbiased worldview can be achieved only if the language that we use is gender-equal (see e.g. Hellinger 1991). This section further examines how language can influence thought in various ways, by considering the concept of 'linguistic relativity'. In particular, the focus lies on ways in which the expression of gender in language may contribute to how we conceptualise the world when we use language.

3.4.1 Overview

The idea that language shapes how we think about reality is not a new one; for instance, the German philosopher and linguist Wilhelm von Humboldt ([1836] 1988: 60) argued as early as 1836 that “[t]here resides in every language a characteristic world-view”. Another influential early figure was the anthropologist and linguist Franz Boas (1858–1942). Boas called into question the idea that some languages and cultures are ‘intellectually inferior’ to others, which was still the prevailing view in the early 20th century. According to Boas, all languages may express the same things, but the tools that they use to do so vary greatly; Boas also maintained that language and culture are inseparable from each other. Nonetheless, in its current shape, linguistic relativity is generally referred to as the Sapir-Whorf hypothesis, named after two central figures: Edward Sapir (1884–1939) and his student Benjamin Lee Whorf (1897–1941). One of Whorf’s (2012 [1956]: 213) main claims was that diversity in language invariably leads to at least some diversity in thought:

The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face; on the contrary, the world is presented in a kaleidoscopic flux of impressions which has to be organized in our minds – and this means largely by the linguistic systems in our minds. We cut nature up, organize it into concepts, and ascribe significances as we do, largely because we are parties to an agreement to organize it in this way – an agreement that holds throughout our speech community and is codified in the patterns of our language.

This assumption does not mean, however, that there are no universal aspects of human cognition.

Although the term ‘Sapir-Whorf hypothesis’ is widely used, it is worth noting that linguistic relativity is not to be understood as a uniform theory, but rather as a group of related hypotheses about the effect language has on thought (for a

review, see Wolff & Holmes 2011). The following syllogism, from Gumperz & Levinson (1996: 24), summarises the core of all relativity hypotheses:

Given that:

(1) differences exist in linguistic categories across languages;

(2) linguistic categories determine aspects of individuals' thinking;

then:

(3) aspects of individuals' thinking differ across linguistic communities according to the language they speak.

The most extreme form of linguistic relativity is 'linguistic determinism', which posits that language determines the way humans think in a fixed, unalterable way. Studies lend no support for a linguistic deterministic view and researchers do not tend to subscribe to this view any longer (see e.g. Lucy 2005). Instead, current research is concerned with more subtle ways in which language and thought interact. Moreover, while early research on linguistic relativity was primarily focused on semantic differences between languages (for an influential study of the colour spectrum, see Berlin & Kay 1969), studies now mainly examine the effects of grammatically encoded differences in languages, such as temporal and numerical representations and formal expression of gender.

One weak manifestation of linguistic relativity is what Slobin (1996; 2003) refers to as 'thinking for speaking'. The main idea behind this framework is that conceptualisations in language steer our thinking when we use language, that is, that language-specific preferences have an effect on *linguistic* thought. Thus, we become accustomed to thinking in certain ways when using language because the structure of our L1 requires us to access certain forms over and over again. For example, studies have shown that the way in which manner and path are encoded in language can affect how individuals recall motion events when asked to describe them. Some languages, including English, encode manner in verbs (e.g. *slip, sneak, stroll*) and use prepositional phrases or particles to encode path (e.g. *into the house*). Other languages, like Spanish, do the opposite, that is, motion verbs express path (e.g. *entrar* 'enter') and optional adverbials encode manner (*entra caminando* '(s/he) enters walking'). It appears that speakers of languages like English are more likely to focus on the manner of a motion event when they describe the event verbally, whereas speakers of languages like Spanish are more likely to attend to the path of the motion (see e.g. Gennari et al. 2002; Papafragou et al. 2008).

The idea of thinking for speaking is in itself not a new one; for instance, Pinker, who has argued strongly against any effects of language on thought (Pinker 1994: 58), wrote the following in 1989 (Pinker 1989: 360):

Whorf was surely wrong when he said that one's language determines how one conceptualizes reality in general. But he was probably correct in a much weaker sense: one's language does determine how one must conceptualize reality when one has to talk about it.

Nonetheless, Slobin (1996) was the first to describe and examine this idea more closely, systematising it as a framework. It is worth noting that the term ‘speaking’ is not necessarily literal as the same kinds of effects should also occur when the language user engages in written modes of language: “the framework embraces all forms of linguistic production (speaking, writing, signing) and reception (listening, reading, viewing), as well as a range of mental processes (understanding, imaging, remembering, etc.)” (Slobin 2003: 160).

Not all researchers regard thinking for speaking as linguistic relativity proper (see e.g. Everett 2013) as the model is restricted to contexts where individuals are actively using their linguistic faculties; for some, claims of relativity effects cannot be made unless there is evidence that language also influences non-linguistic thought processes, in addition to linguistic thought. However, studies have suggested that a meaningful distinction is, in fact, difficult to make between linguistic and non-linguistic thought: “What researchers have been calling ‘thinking’ this whole time actually appears to be a collection of both linguistic and nonlinguistic processes. As a result, there may not be a lot of adult human thinking where language does not play a role” (Boroditsky 2011: 65). Moreover, even if the two could be clearly separated, linguistic thinking still constitutes a substantial part of human cognition:

Humans use language incessantly, for some large (though variable and admittedly undetermined) portion of their waking hours. So even if linguistic effects on cognition are restricted to the realm of language, a point that is vociferously disputed by proponents of linguistic relativity, it is worth stressing what a large realm that is! (Everett 2013: 32)

If one goes back to the roots of linguistic relativity, the pioneers Boas, Sapir, and Whorf were, in fact, primarily interested in verbal performance and linguistic thought, in a similar fashion to thinking for speaking (see e.g. Jarvis 2016; Pavlenko 2014). The view taken in this dissertation is that the framework of thinking for speaking is to be understood as a weak form of linguistic relativity, in line with, for instance, Wolff & Holmes (2011) and the studies in Han & Cadierno (2010).

3.4.2 Previous research on grammatical gender and linguistic relativity

A plethora of studies have examined the influence of grammatical gender on thought (for a review, see Samuel et al. 2019). For instance, Boroditsky et al. (2003) investigated how grammatical gender marking might affect conceptualisations of inanimate (non-sexed) entities. In this study, German and Spanish speakers were asked to describe masculine and feminine inanimate nouns using adjectives. The study found that the participants used more stereotypically female adjectives when the noun was feminine (e.g. *little* and *intricate* in describing the Spanish noun *llave* ‘key (f)’), and more stereotypically

male adjectives when the noun was masculine (e.g. *hard* and *heavy* in describing the German noun *Schlüssel* 'key (m)'). However, these findings could not be replicated in a later study by Mickan et al. (2014). Research has also suggested that potential relativity effects are stronger in languages that only encode two genders (masculine and feminine), like French and Italian, than in languages that have three genders (masculine, feminine, and neuter), like German (see e.g. Sera et al. 2003). A plausible reason for this difference is that the grammatical gender of nouns map more easily onto biological gender when the gender system only has two categories (see e.g. Vigliocco 2005).

Studies have also examined how gender marking influences conceptualisations when the target is a human entity, that is when biological sex is, in fact, a relevant feature. The findings of these studies indicate that relativity effects are stronger for human entities than for non-human ones (see Everett 2012; Ramos & Roberson 2011; Samuel et al. 2019; Vigliocco et al. 2005). It is likely to be this way because there is a clear conceptual connection between grammatical gender and biological sex when the referent is animate, particularly so with human referents.

Relativity effects have also been found for languages with natural gender, like English, where the formal expression of gender is limited to human referents, in the form of personal pronouns. In a study by Chen and Su (2011), L1 speakers of Chinese (a genderless language) and L1 speakers of English (a language with natural gender) were asked to answer gendered and non-gendered questions after hearing a story and after matching pictures with sentences. The results showed that the English-speaking participants were significantly faster and more accurate at responding to the gender-related questions than the Chinese speakers. As argued in the study, these findings suggest that English speakers have a higher sensitivity to gender because of the existence of natural gender in their L1. Another study compared speakers of English and speakers of Karitiâna, a language that does not have gender marking in the pronoun system (Everett 2011). In this study, the participants watched video clips that featured gender-ambiguous faces and they were then asked to describe what they had seen in the videos. The results showed that the English speakers were more prone to give androcentric responses to the stimuli than the Karitiâna speakers. These findings suggest that the English speakers were influenced by the pronominal distinction in English, and, more specifically, the traditional use of *he* as a generic pronoun.

3.4.3 Relating linguistic relativity and L2 use of language

Studies of linguistic relativity in the 1980s and the early 1990s primarily focused on relativity effects in the L1, either including only monolingual language users or ignoring the possible influence on any additional languages that the users may speak (see e.g. Pavlenko 2014). Thus, these studies failed to acknowledge the fact that, globally speaking, multilingualism is the norm and monolingualism the exception. In recent research, however, the scope has been broadened to include speakers of multiple languages. The framework of thinking for speaking is now

also applied to the use of L2 languages in addition to L1 use (for a study on thinking for speaking in L2 written English, see Ekiert 2010; see also Pavlenko 2014). Slobin (1996: 89) describes the interaction between the L1 and the L2(s) in the following way:

In brief, each native language has trained its speakers to pay different kinds of attention to events and experiences when talking about them. This training is carried out in childhood and is exceptionally resistant to restructuring in adult second-language acquisition.

Based on this hypothesis, speakers of languages with formal expressions of gender have been 'trained' to be attentive to gender when using language in general, which may also be reflected in L2 use. An example of the effects of L1 grammatical gender on L2 processing is given in Boroditsky et al. (2003): in this study, L1 Spanish speakers and L1 German speakers were asked to remember proper names for 24 objects in English; for example, the object 'apple' was named either *Patrick* or *Patricia*. The findings showed that both Spanish and German speakers were better at remembering object-name pairs when the grammatical gender of the noun in their L1 matched the gender of the proper name that the object was assigned in English.

The influence of L1 conceptualisation patterns may take different forms in L2 use. A very common phenomenon in L2 acquisition is crosslinguistic influence (CLI), which refers to ways in which the L1 can influence the use of the L2 (see e.g. Ringbom 2007). Crosslinguistic influence includes transference of lexical items, but it also affects many other domains of language, such as syntax, pragmatics, and phonology. A specific type of crosslinguistic influence is 'conceptual transfer', which refers to interaction between the L1 and the L2 where relativity is also involved (see e.g. Jarvis 2016; Odlin 2005). Relating conceptual transfer to Slobin's (1996) framework, Pavlenko (2011: 246) describes this phenomenon as "thinking in L1 for speaking in L2". Research has shown that grammatical gender is part of the conceptual system (see e.g. Jarvis & Pavlenko 2008): for instance, all gender forms are stored in the language user's mental lexicon, and gender information is automatically activated when language users are presented with nouns that carry gender (for a review, see Schriefers & Jescheniak 1999). Jarvis & Pavlenko (2008: 136) distinguish three main ways in which conceptual transfer may manifest itself in the domain of gender:

in cases where L2 learners fail to attribute grammatical gender to all or most entities, in cases where particular mental representations fail to encode [GENDER], and in cases where mental representations transfer the [GENDER] component from one language into the other

For the discussion of L2 English, the latter two are of relevance as English does not have grammatical gender. The extent to which conceptual transfer occurs

will depend on individual factors such as frequency of language use, language dominance as well as language proficiency (see e.g. Pavlenko 2014). For instance, conceptual transfer tends to be more frequent in the beginning stages of language acquisition than in advanced learning. The same individual factors may also influence to what degree cognitive restructuring is likely to take place, that is, when L2 learning leads to attainment of a new conceptual contrast or to suppression of a salient L1 contrast (see e.g. Pavlenko 2011, 2014). Prolonged exposure to an L2 may even lead to the reversed kind of conceptual transfer, where the L2 influences L1 conceptualisation (see e.g. Bylund & Jarvis 2011).

4 Epicene pronouns in English

The only elements that show gender agreement with nouns in English are personal pronouns in the third-person singular. The choice of personal pronouns is relatively straightforward in referential use as it is based on the biological gender, or sex, of the referent. However, choosing an appropriate pronoun becomes problematic when the antecedent is a human being of indeterminate gender. As stated in the introduction, personal pronouns used with this type of antecedent are called epicene pronouns. This chapter gives a detailed description of epicene pronouns in English, including the types of antecedents with which these pronouns co-occur. The chapter ends with a discussion of previous research on the use of epicene pronouns in different varieties of English.

4.1 *He, he or she, or they?*

The issue of epicene pronouns has featured prominently in discussions about sexist language use in English. Some researchers have referred to this problem as a ‘gap’ in the pronominal system (see e.g. Weidmann 1984), and there have been many attempts at introducing a new, epicene pronoun in English. Examples of such neologisms include the pronouns *thon* and *ze* (for an overview, see Baron 1986: 205-208). Thus far, however, none of these innovations has reached widespread adoption in a similar way as the pronoun *hen* in Swedish (see 3.3). As suggested by Everett (2013: 241), one reason for the failed attempts is that the prevailing system is not necessarily considered flawed:

Part of the motivation for this failure is that many grammarians have overtly questioned the possibility that reliance on a gendered 3rd person pronoun actually impacts the manner in which people think of human referents. Attempts at epicene pronoun introduction have been motivated by the opposite intuition, namely that gendered 3rd person pronouns cause people to think in gendered ways, even when gender may be irrelevant to a given 3rd person reference.

The justifications mentioned in this quote are reminiscent of linguistic relativity and the idea that language affects the way we think. However, the global nature of the English language is also likely to factor in here; for a new pronoun to become widespread, an immense number of people would have to accept and adopt the innovation.

The most frequently used and discussed types of epicene pronouns in present-day English are generic *he*, combined forms like *he or she* and *s/he*, and singular *they* (see e.g. Baron 2020; Paterson 2014). Although less common, the pronouns *one* and *she* are also sometimes used in generic reference in English (see e.g. Huddleston & Pullum 2002; Quirk et al. 1985). This dissertation is concerned with the use of *he*, *he or she* variants and *they* as other pronoun types

proved to be so rare in the data (see 5.1.2; see also Laitinen 2007; Paterson 2014).

Diachronically speaking, the pronoun *he* acquired the function of an epicene pronoun because the generic word for person in Old English, *mann*, belonged to the masculine gender (see e.g. Aikhenvald 2016: 195). Nonetheless, *he* continued to be used in a generic way in Middle English, despite the loss of grammatical gender. This convention was viable in the sense that men were literate to a much larger extent than women, and texts were generally written for men, about men. Yet, different types of epicene pronouns existed alongside each other, and prescription of the use of these pronouns of English did not start until the end of the 18th century, which was an intense period in terms of codification and standardization of the language. From this point until as late as the 1970s, grammar books and style guides prescribed the use of generic *he* for singular gender-indefinite reference in English. In current times, however, it would be rare to find a grammar book that recommends generic *he* (see e.g. Paterson 2014).

It has been argued that the use of generic *he* is simply a matter of grammatical convention, separate from any social constructions of gender (for a discussion, see e.g. Martyna 1980: 483f). However, a plethora of studies have shown that generic *he* does not function as a neutral pronoun in English as the use of *he* tends to be interpreted as referring to men rather than individuals of any gender identity (see e.g. Gastil 1990; Hamilton 1988; MacKay & Fulkerson 1979; Miller and James 2009; Wilson & Ng 1988; for a review of research on the comprehension of generic masculines in English, see Henley & Abueg 2003). It can thus be concluded that the use of generic *he* “adds the property of maleness to a generic discourse referent” (Balhorn 2004: 84). This finding is similar to what research has revealed about the use of generic masculines in languages with pervasive grammatical gender (see 3.3).

As the use of generic *he* fell out of favour, something else needed to take its place. A commonly proposed strategy for avoiding gender-bias is the use of coordinations and composites, like *he or she* and *he/she*. The use of these forms is best understood as a feminisation strategy in English as such use increases female visibility, rather than neutralises gender altogether. Early grammarians, however, dismissed *he or she* variants on the basis that they are cumbersome and unwieldy, and this argument against such forms is still found to some degree in grammars and style guides (see e.g. Baron 2020; Huddleston & Pullum 2002; Quirk et al. 1985). Moreover, studies have suggested that the ordering of the pronouns carries meaning and thus affects how coordinations and composites are interpreted: when the male pronoun precedes the female one – the most common order of the pronouns – there seems to be a male bias in individuals’ mental representations (see e.g. Gastil 1990: 640). In recent times, another issue that has been highlighted in discussion is that *he or she* variants sustain a binary view of gender, thus excluding individuals who do not identify within the male/female binary (see e.g. Bradley et al. 2019).

Another alternative to the use of generic *he* is singular *they*. The use of *they* as a singular pronoun emerged as a consequence of the loss of grammatical gender in English, that is, this use was not initially related to any socio-cultural factors or a strive for 'political correctness' (Balhorn 2004). However, as with *he or she* variants, the increase in the use of singular *they* since the 1970s is largely a consequence of feminist language reform. The use of singular *they* can be described as a neutralisation strategy in English since the mentioning of gender is completely avoided by using this pronoun. Thus, singular *they* is arguably the only truly gender-neutral epicene pronoun in current English use, as *he or she* variants are still gendered.

The use of singular *they* is not in any way a new phenomenon in English; there is evidence to suggest that this use of the pronoun may have existed for as long as 400 years (see e.g. Balhorn 2004; Curzan 2003). Despite this fact, the use of *they* with singular referents was condemned by grammarians for at least a century, starting from the end of the 18th century (Curzan 2003: 73). The main argument against singular *they* is that its use purportedly violates number-agreement rules in English (see e.g. Huddleston & Pullum 2002; Quirk et al. 1985). However, this is only an issue insofar as singular *they* is construed as a plural pronoun; as argued by Paterson (2014), there is strong evidence to suggest that a majority of L1 speakers of English perceive singular *they* as a singular pronoun proper. For instance, singular *they* sometimes occurs in referential use when the sex of the referent is irrelevant, as exemplified in (7) from Balhorn (2004: 84). In cases like this, the use of *they* cannot be accounted for by notional plurality of the antecedent.

(7) Somebody called while you were out and *they* said *they*'d call back later.

Although the verb following singular *they* is always in the same form as for the third-person plural, this is not necessary grounds for claiming that the pronoun cannot be singular – the same is also true of singular *you*, which was, in fact, also proscribed in its early days of use (see e.g. Baron 2020). Historically speaking, the most accepted use of singular *they* has been with notionally plural antecedents, such as *everyone* (see e.g. Meyers 1993); grammars have typically included examples of singular *they* referencing antecedents of this kind (Meyers 1993; Paterson 2014).

In recent years, singular *they* has also taken on the function of a non-binary pronoun, that is a pronoun used in reference to individuals who do not identify as either male or female (see e.g. Baron 2020; Bradley 2020). Although various neologisms have been suggested for non-binary use, such as *ze* and *thon*, research suggests that *they* is currently the most used and widely accepted non-binary pronoun (see e.g. Baron 2020; Bradley et al. 2019; Hekanaho 2020). It is beyond the scope of this dissertation to further examine non-binary uses of singular *they*.

4.2 Epicene antecedents

Epicene pronouns are used in co-reference with different kinds of antecedents. These antecedents can be categorised into two main groups: (1) singular indefinite pronouns, like *anyone* and *somebody*, and (2) singular NPs headed by a gender-indefinite noun, like *a child* or *any person*. The second of these groups, NPs, can be further divided into three subtypes: (1) definite NPs, such as *the student*, (2) indefinite NPs, such as *a student*, and (3) quantificational NPs, such as *no student* (see e.g. Paterson 2014).

As the categorisation suggests, epicene antecedents are highly variable in their nature. For instance, antecedents differ with respect to their notional or semantic number, that is, whether they have a singular or a plural meaning. Table 2 shows the different types of antecedents as well as their notional number.

Table 2. Types of epicene antecedents and their notional number

Type	Form	Notional number
Indefinite pronouns	<i>everyone/everybody</i>	plural
	<i>anyone/anybody</i>	singular/plural
	<i>no one/nobody</i>	singular/plural
	<i>someone/somebody</i>	singular
Quantificational NPs	<i>every</i> <noun sg>	plural
	<i>each</i> <noun sg>	plural
	<i>any</i> <noun sg>	singular/plural
	<i>no</i> <noun sg>	singular/plural
	<i>some</i> <noun sg>	singular
Indefinite NPs	<i>a</i> <noun sg>	singular
	<i>an</i> <noun sg>	singular
Definite NPs	<i>the</i> <noun sg>	singular
	<i>this</i> <noun sg>	singular
	<i>that</i> <noun sg>	singular

In terms of notional number, antecedents including the quantifiers *every* and *each*, as in *everyone* and *each student*, are inherently plural in their meaning. In contrast, antecedents including the quantifier *some* are notionally singular, as are definite and indefinite NPs. Some antecedents, like *no one* and *any student*, can be either notionally singular, notionally plural, or somewhere in between, depending on the context.

Epicene antecedents can also be categorised according to their ‘degree of individuation’. The concept of individuation describes how easily an antecedent may be construed as a specific entity (see e.g. Fraurud 1996; Hopper &

Thompson 1980; Timberlake 1977). Timberlake (1977: 162) lists five aspects of individuation, which are illustrated in Table 3. In addition to these, Hopper & Thompson (1980: 253) include the aspect of referentiality, which also appears in the Table. In all cases, the feature to the left is more individuated than the one to the right.

Table 3. Aspects of individuation

High individuation	Low individuation
	proper > common
	human > animate > inanimate
	concrete > abstract
	singular > plural
	definite > indefinite
	referential > non-referential

For the discussion of epicene antecedents, two of these features are particularly relevant: the distinction between singular and plural, and the distinction between definite and indefinite. As concluded above, epicene antecedents may have a plural meaning despite their grammatical singularity. Thus, notionally plural antecedents, like *everyone*, are less individuated than notionally singular ones, like *someone*. Because of the plural meaning, the boundaries are fuzzier and the notionally plural antecedent is thus less likely to be conceptualised as a specific individual. Moreover, indefinite antecedents, like *a child*, are less individuated than definite ones, like *the child*, which denote a specific human entity.

A phenomenon that relates to individuation is social gender or gender expectancy. While indefinite pronouns are devoid of gender connotations, head nouns in NP antecedents differ in the degree to which they are associated with male or female referents. Several studies have identified the stereotype value of personal nouns, both in English (Abudaljuh 2012; Carreiras et al. 1996; Doherty & Conklin 2017; and Kennison & Trofe 2003), and in other languages (Abudaljuh 2012; Misersky et al. 2014). These studies have found, for instance, that the nouns *pilot* and *farmer* are strongly associated with male individuals in English, whereas *housekeeper* and *babysitter* are associated with women. Examples of neutral nouns include *student* and *child*. It should be noted that while an antecedent in itself may be free of gender associations, contextual and discursive factors can still lead to stereotyping. For instance, if a speaker says that they met *a nice person in the beauty salon*, their interlocutor may draw the conclusion that this person is a woman even though the antecedent *person* is neutral.

Although gender stereotyping is distinct from individuation, the two notions bear similarities. A highly individuated antecedent is easy to conceptualise as a specific entity (e.g. *this man*), whereas this is not true of antecedents that are low

in individuation (e.g. *anyone*). In a similar fashion, an antecedent with high gender expectancy is likely to evoke a more specific mental image of the referent in the language user's mind, than an antecedent that is free of gender connotations.

4.3 Research on the use of and attitudes to epicene pronouns

The use of epicene pronouns is well-researched in inner-circle varieties of English, both in spoken registers (see e.g. Gerner 2000; Matossian 1997; Newman 1992, 1998; Pauwels 2001; Stringer & Hopper 1998) and in written registers (see e.g. Baranowski 2002; Laitinen 2007; Meyers 1990; Paterson 2011, 2014, 2020; Strahan 2008). Yet, in contrast, only limited attention has so far been given to epicene pronouns in learner varieties of English, as well as in English as a lingua franca (but see Abudalbuh 2012; Lee 2007). This section presents a summary of key findings of previous research on both L1 and L2 English use.

A number of previous studies of L1 English have reported diachronic changes in the use of epicene pronouns (see e.g. Cooper 1984; Earp 2012; Hegarty & Buechel 2006; Laine & Watson 2014; Laitinen 2007; Paterson 2011, 2020; Pauwels 2001). One of the most striking changes is the rapid decline in the use of generic *he* since the 1970s and the second wave of feminism; such a tendency has been found, for instance, in British and American newspapers and periodicals (Cooper 1984; Laine & Watson 2014; Paterson 2011), in published journal articles (Hegarty & Buechel 2006), and in Australian radio programmes (Pauwels 2001). While the use of generic *he* has decreased, the use of singular *they* has, in turn, become considerably more frequent in recent decades. Again, this development has been observed across genres and registers (see e.g. Laitinen 2007; Paterson 2011). There is also evidence of a diachronic change in the frequency of *he or she* variants: in a study by Curzan (2014), the use of *he or she* forms increased considerably from the 1970s until the 2000s in spoken and written American English. However, it still appears that *he or she* variants have not grown in popularity to the same degree as singular *they* (Paterson 2020), and the most recent studies of L1 English thus suggest that *they* is the preferred epicene pronoun in inner-circle varieties (see e.g. Hekanaho 2020; Laitinen 2007; Paterson 2014).

The small number of studies of L2 use suggest that L2 learners may follow different patterns from inner-circle varieties. In a study by Abudalbuh (2012), 100 L1 Arabic university students completed a number of language tasks that were designed to elicit the use of epicene pronouns. The findings showed that the most frequent pronoun was generic *he* (71%, $N = 2,194$), rather than singular *they* (6%) or *he or she* variants (8%). A similar study was conducted by Lee (2007), who examined the use of epicene pronouns in a set of elicitation tests that were administered to 269 L1 Chinese (Cantonese) students. In this study,

too, generic *he* proved to be the most frequently used epicene. Furthermore, a survey reported in Pauwels (2011: 16) showed that L2 speakers of English believe that the strategies available in their L1, such as feminisation, are transferable to other languages, and that L2 speakers claim to draw on L1 strategies for achieving gender-inclusive language in English.

Both L1 and L2 research has found systematic variation in the use of epicene pronouns that can be tied to the co-referential antecedent. In particular, notional number appears to be a strong predictor for the frequency of the different types of epicenes: while singular *they* is more likely to be used with notionally plural antecedents than with notionally singular ones, generic *he* and *he or she* variants show the opposite pattern of distribution. These tendencies have been observed in both L1 use (see e.g. Baranowski 2002; Gerner 2000; Hekanaho 2020; Newman 1992, 1998; Paterson 2014) and L2 use (Lee 2007). Another relevant aspect of the antecedent is social gender. Studies have disclosed that generic *he* is more likely to be used in reference to antecedents that are high in male gender-expectancy, than in reference to antecedents that are neutral or stereotypically female (Abudaljuh 2012; Pauwels 2001).

The choice of epicene pronouns may also be affected by the degree of formality of the context (see e.g. Baranowski 2002; Hekanaho 2020; Holmes 1998; Parini 2012). For example, Curzan's (2014) study of spoken and written American English revealed that the use of *he or she* variants was much more common in formal texts than in fiction writing and in spoken language. In turn, Holmes (1998) studied epicene pronouns in the *Wellington Corpus of Spoken New Zealand English* (WCSNZE), which includes both formal and informal speech. The results showed that singular *they* was considerably more frequent in informal speech than in formal speech, while generic *he* showed the opposite kind of distribution.

The difference in formality between the pronoun types is also visible in language users' attitudes. For instance, studies have revealed that the use of singular *they* is still viewed with some suspicion by L1 speakers of English (Bradley et al. 2019; Hekanaho 2020; LaScotte 2016; Madson & Hessling 2001) although this situation may change rapidly as formal endorsement of the pronoun increases (see e.g. *Publication Manual of the American Psychological Association* 2019). In a rating study conducted by Madson & Hessling (2001), American university students rated texts containing singular *they* as "lower in overall quality" compared with texts containing *he or she* or generic *he* and generic *she* alternately. In another two recent studies, *he or she* variants were still rated as more grammatical than singular *they* although generic *he* was viewed as the least acceptable solution out of the three pronoun types (Bradley et al. 2019; LaScotte 2016). As far as L2 users are concerned, there is some evidence that they, too, have a perception of singular *they* as non-standard or incorrect (see Hekanaho 2020; Pauwels 2011). In the survey reported in Pauwels (2011: 16), L2 participants expressed concerns about losing linguistic credibility by using singular *they* in place of generic *he*. The participants' main worry was that other language users would regard them as 'non-competent

speakers' for using *they*, rather than individuals who want to express themselves in a non-biased way.

5 Materials and methods

This chapter presents the materials and methods that are used in the four substudies. Three of the studies are corpus-based analyses of the use of epicene pronouns, while one study is based on data elicited from an experiment. All the studies combine quantitative and qualitative analytical techniques. Table 4 gives an overview of the types of data, the groups of language users and the variables that form the foundation of each study.

Section 5.1 presents the materials and methods used in the three corpus-based studies while section 5.2 focuses on the experimental study. Section 5.3 presents the methods of analysis.

Table 4. Details of the four substudies

Study	Types of data	Groups	Variables
Study 1	Argumentative student essays	<ul style="list-style-type: none"> • L1 speakers of English • L2 speakers of English from 13 L1 backgrounds 	<ul style="list-style-type: none"> • L1 background • Antecedent type
Study 2	Data elicited from an experiment; the language in the tasks represents media discourse aimed at the general public	L2 speakers of English from 8 L1 backgrounds	<ul style="list-style-type: none"> • L1 background • Antecedent type • Gender expectancy of the antecedent
Study 3	Student academic writing: term papers and BA and MA theses	L2 speakers of English with Swedish as L1	<ul style="list-style-type: none"> • Decade when the text was produced • Antecedent type • Gender expectancy of the antecedent
Study 4	Research articles published in open access journals; author guidelines	<ul style="list-style-type: none"> • Researchers of languages and linguistics • Researchers of library and information science 	<ul style="list-style-type: none"> • Research field • Antecedent type • Gender expectancy of the antecedent

5.1 Corpus linguistics

Corpus linguistics has become a widely used methodology in linguistic research, not least in applied linguistics. The first learner corpora emerged in the 1980s, and corpora are now frequently used to examine learner language at various stages (see e.g. Granger 2015). The development in corpus research is in many ways the result of major advances in technology, which have made it possible for researchers to store and analyse very large data sets. As outlined in Biber, Conrad & Reppen (1998: 4), linguistic analyses using corpora are characterised by four essential aspects: (1) they examine natural texts that have been compiled into corpora in systematic ways, (2) they examine real patterns of use in an empirical way, (3) they make use of computerised methods, applying both manual and automated techniques, and (4) they include both quantitative and qualitative procedures.

A distinction can be made between corpus-based, corpus-driven, and corpus-assisted approaches (see e.g. Baker 2010; Tognini-Bonelli 2001). Corpus-based approaches start from existing hypotheses about language phenomena, that is, the researcher uses corpora mainly for testing out existing ideas and assumptions. While corpus-based approaches are thus deductive in nature, corpus-driven approaches are inductive. This means that the researcher uses corpus-linguistic methods to explore patterns without any strong assumptions about what these patterns might be. Corpus-assisted approaches, finally, tend to be abductive as the researcher sets off with certain hypotheses but also explores new patterns in the corpus. Corpus-assisted approaches normally supplement corpus linguistics with other types of methods, in order to examine the corpus materials from different angles. For instance, corpus-assisted studies may include interviews and etymological approaches (see e.g. Baker 2010). The present dissertation takes on a corpus-based approach, which, in addition to corpus-linguistic techniques (Studies 1, 3, 4), also includes experimental methods (Study 2).

Corpus-linguistic methods include a variety of processes and techniques, such as measuring frequency, collocation, and dispersion (Baker 2010: 19). The substudies in this dissertation are concerned with the frequency of different types of epicene pronouns. By measuring frequencies, it is possible to establish and assess synchronic as well as diachronic trends in the use of pronouns. Each corpus study examines the distribution of the three most common types of pronouns: generic *he*, *he or she* variants, and singular *they*. The distributions are measured according to four main parameters: (a) L1 group (Study 1), (b) the antecedent type and gender-expectancy of the antecedent (Studies 1, 3, 4), (c) academic discipline (Study 4), and (d) the decade when the text was produced (Study 3); consider Table 4 above.

A major benefit of using corpus linguistics is that a large collection of texts can be examined for a given linguistic phenomenon, which is necessary when measuring frequencies. With respect to the present dissertation, a total of 4,921 instances of epicene pronouns were extracted in the three corpus-based studies. Such extensive data would have been difficult to extract and analyse using other

techniques; a solid set of pronoun data including textual context was needed in order to examine trends in the use of epicene pronouns in different groups of EFL learners and ELF users.

Another oft-cited advantage of corpus linguistics is that it focuses on naturally occurring language, in contrast to language data that have been collected using elicitation techniques (see e.g. Gilquin & Gries 2009; see also 5.2). However, learner corpus materials (Study 1) still tend to be elicited to some degree; for instance, participants may be asked to produce text on a topic that has been given to them by the researcher, and there may also be limits on how much time the participants can spend on the task at hand. Thus, the 'naturalness' of written corpus materials may differ to a fair degree.

As far as academic corpora are concerned, it is also worth noting that the texts are rarely the sole production of one single individual, but they reflect a collective notion of authorship. This situation is true of both student academic writing and published academic articles (Studies 3-4). For some types of student texts, such as theses and dissertations, supervisors may give comments on drafts of the texts and students will make changes in accordance with such suggestions. Academic research articles, in turn, tend to be the result of thorough editing, and both editors and reviewers may suggest changes that have bearing on the use of language in a text.

5.1.1 Corpora

A total of five corpora of written English were used in Studies 1, 3, and 4 to collect data on the use of epicene pronouns. The specifics of each corpus are shown in Table 5, with a more detailed description of each of the corpora following below.

Table 5. Corpora used in the dissertation

Corpus	No. of words	Time period	Types of texts	L1 background	Study
ICLE	ca. 2.8 million	1991-2004	Student argumentative essays and literature examination papers	Bulgarian, Czech, Dutch, Finnish, French, German, Italian, Norwegian, Polish, Russian, Spanish, Swedish, Turkish	1
LOCNESS	ca. 320,000	1991-1995	Student argumentative essays and literature examination papers	English	1
BATMAT	ca. 2.5 million	1972-2016	Student academic papers	(Finland-)Swedish	3
VESPA-SE	ca. 280,000	2014	Student academic papers	Swedish	3
COAJA	ca. 6 million	2017-2018	Academic journal articles and author guidelines	Various	4

The *International Corpus of Learner English* (ICLE v2, see Granger et al. 2009) comprises argumentative essays and literature examination papers written by university students of English across the world; in total, sixteen L1 backgrounds are represented in the corpus. For the purpose of Study 1, the thirteen European L1 subcorpora of the ICLE were used: Bulgarian, Czech, Dutch, Finnish, French, German, Italian, Norwegian, Polish, Russian, Spanish, Swedish, and Turkish. These subcorpora were collected between 1991 and 2004 and include 4,218 essays totalling approximately 2.8 million words of running text. The basis for using the ICLE was that this learner corpus is unprecedented in its scope, including such a large variety of L1 groups for comparison. As the goal of the dissertation was to compare the findings of each study, the choice of learner groups in the subsequent substudies was guided by the fact that the ICLE includes university students of English Language and Literature.

The *Louvain Corpus of Native English Essays* (LOCNESS, see Centre for English Corpus Linguistics 2016) was designed specifically to be comparable to the ICLE. The essays in LOCNESS are therefore of a similar kind as the texts in the ICLE, but the writers are British and American L1 speakers of English. The British writers include both A-level students and university students, whereas the American writers are all university students. In total, the corpus comprises 436 texts or 324,304 words of running text.

The BATMAT corpus (Lindgrén 2016) is a diachronic corpus of student academic writing, which consists of Bachelor's (BA) and Master's (MA) theses written by students of English Language and Literature at Åbo Akademi University in Turku, Finland. The texts were produced between 1972 and 2016. In total, the corpus comprises 121 theses or approximately 2.5 million words. The vast majority of the student writers in the corpus are Finland-Swedes, meaning that their L1 (or one of their L1s) is Swedish.

Like the BATMAT corpus, the *Swedish subset of the Varieties of English for Specific Purposes dAtabase* (VESPA-SE, see Larsson 2015) is a corpus of student academic writing. VESPA-SE includes 69 term papers and BA theses – ca. 280,000 words – written by students of English Language and Literature at Uppsala University and Stockholm University in Sweden. All texts were produced in 2014, and the writers are L1 speakers of Swedish.

The *Corpus of Open Access Journal Articles* (COAJA) was compiled by the author in late 2018 and early 2019, and it consists of academic papers published in 20 open access (OA) journals in 2017 and 2018. The articles represent two academic research fields: (1) languages and linguistics, and (2) library and information science. In total, the corpus comprises 1,003 articles or ca. 6 million words. Because of the nature of these corpus materials, the L1 backgrounds of the writers are not known; nonetheless, it can be assumed that the writers – researchers publishing in international journals – include both L1 and L2 speakers of English, from a wide range of lingua-cultural backgrounds. The corpus also includes the author guidelines for each of the 20 OA journals, as well as for 20 subscription-based journals. The non-OA journals were included for comparison and they were selected from the SCImago Journal Index (SCImago

Lab 2018), which ranks journals according to citations. The SCImago Journal Rank (SJR) indicator is used as an alternative to impact factor and is based on data from the Scopus database.

5.1.2 Extracting epicene pronouns

The three corpus-based studies, Studies 1, 3, and 4, all followed the same procedures for extracting epicene pronouns. The corpus searches were conducted using the concord function in WordSmith Tools 5.0 (Scott 2008). Table 6 shows the types of search terms that were employed in WordSmith Tools. The search words consisted of (1) eight indefinite pronouns: *everyone*, *everybody*, *anyone*, *anybody*, *no one*, *nobody*, *someone*, and *somebody*; and (2) ten determiners: *any*, *every*, *each*, *no*, *some*, *a*, *an*, *the*, *this*, and *that*. The purpose of searching for determiners was to find constructions with epicene pronouns used in co-occurrence with NPs. Searching for determiners also allowed for lexical elements to occur between the determiner and the noun in the NP antecedent, as in *every deaf child* and *the most committed IL developer* (examples from the data for Study 4). As context words, all forms of generic *he*, *he or she* variants, and singular *they* were searched for. The context search horizon was set to 0 words left of the search word and 25 words right of the search word, which means that the pronoun (context word) had to occur within the 25 words following the antecedent (search word). Although pronouns could potentially occur further away from the antecedent, it was expected that the large majority of pronoun instances would fall within the limit of 25 words (for a similar approach, see Gerner 2000: 96; Laitinen 2007: 109).

In preparation for the first study, pilot searches were conducted with only indefinite pronouns as search words, leaving out the context words. This procedure was carried out in order to explore the possibility that pronouns other than *he*, *he or she* variants and singular *they* might be in frequent use. Indefinite pronouns were chosen for the approach because they give a considerably more limited set of search results than searching for determiners without any context words would do. Nonetheless, no other frequent pronouns were found using this approach. In all three studies, preliminary searches were also conducted for the pronoun generic *she*, but as no or very few instances were found in each case, *she* was not further examined in any of the studies. In the small number of instances where *she* occurred, the context made it clear that the intended referent was a woman rather than an individual of indeterminate gender, as seen in example (8) below.

- (8) So, when *someone* has abortion *she* kills *her* baby on purpose and *she* becomes a murderess. (ICLE, Study 1)

All instances of pronouns were examined carefully, to make sure manually that they were, in fact, occurrences of epicene pronouns and not referential uses of personal pronouns. This approach included examining the co-text of each instance.

Table 6. Search words and context words in WordSmith Tools

Search words	Context words	Context search horizon
Indefinite pronouns <i>everyone, everybody</i> <i>anyone, anybody</i> <i>no one, nobody</i> <i>someone, somebody</i>	<i>he, his, him, himself</i> <i>he or she, his or her,</i> <i>him or her, himself or</i> <i>herself, him- or herself</i>	0 words left of the search word 25 words right of the search word
Determiners <i>any, every, each, no,</i> <i>some</i> <i>a, an</i> <i>the, this, that</i>	<i>she or he, her or his,</i> <i>her or him, herself or</i> <i>himself, her- or himself</i> <i>he/she, his/her,</i> <i>him/her,</i> <i>himself/herself</i> <i>she/he, her/his,</i> <i>her/him,</i> <i>herself/himself</i> <i>s/he</i> <i>they, their, them,</i> <i>themselves, themselves</i>	

In all three studies, condensed tokens were used in establishing the frequency of occurrences. This means that only the first instance of an epicene pronoun with a given antecedent was counted rather than counting all occurrences of pronouns with each antecedent (for discussions of using condensed tokens, see Newman 1992: 456; Paterson 2014: 53). Thus, the four uses of *he/she* in example (9) would be counted as one instance only of *he or she* variants.

- (9) *The person* also evaluates the attributions made to and the construals of others and *his/her* own self (evaluation), which may then result in changes to *his/her* identity, that is, how *he/she* construes *himself/herself*. (COAJA, Study 4)

When more than one type of pronoun co-occurred with one and the same antecedent, one instance each of the given pronoun type was included in the data sets. The occurrences of pronouns in example (10) would therefore be counted as one instance of *he or she* forms and one instance of singular *they*.

- (10) Their belief is that God sees all the actions of people and that *everyone* will be judged according to what *he or she* has done in *their* life. (ICLE, Study 1)

A major benefit of using condensed tokens is that multiple references to individual antecedents do not skew the figures.

5.2 Experimental techniques

As Ellis & Barkhuizen (2005: 49) aptly point out, “no one method will provide an entirely valid picture of what a learner knows or thinks”. Thus, applying a multi-method approach can be beneficial in gaining a deeper understanding of L2 language use (see e.g. Gilquin 2007). In addition to corpus linguistics, experimental methods are often employed to explore learner language. Such methods include a wide variety of techniques, ranging from elicitation of data to tracking of eye-movements (for an overview, see Gilquin & Gries 2009). With respect to this dissertation, the data for Study 2 were collected using an online elicitation experiment. Corpus linguistics and experimental methods tend to complement each other in terms of their advantages and disadvantages, which make them highly suitable to combine in research (see Gilquin & Gries 2009). A major advantage of experimental techniques is that they can be tailored to the specific questions that the research focuses on. This is particularly useful if the studied phenomenon is relatively rare as it may not occur frequently enough in corpus materials to allow generalisations about its use in a given learner group (see e.g. Gilquin & Gries 2009). For the present work, a benefit of applying elicitation techniques was that the use of epicene pronouns could be tested with a specific set of antecedents which remained the same across L2 groups. Such precise comparisons are difficult to make using corpus linguistic methods as the same individual antecedents are unlikely to be used by each student in the corpus materials. Moreover, it was possible to include antecedents in the experiment that are usually rather infrequent in written corpus materials, such as highly stereotyped antecedents (see e.g. Paterson 2014).

A drawback of experimental methods is that elicited language use tends to be less natural than that found in corpus data (see e.g. Gilquin & Gries 2009; see also Section 5.1) in the sense that the experimental design and the experimental setting will influence how participants express themselves. For example, they may modify their language use to fit the perceived purpose of the experiment as well as the expectations of the researcher. To mitigate these issues, the experiment can be performed in a setting that – to the degree that this is possible – does not remind the participants of a ‘school test situation’. In terms of the actual design of the experiment, filler items are useful for drawing the participants’ attention away from the linguistic phenomenon in question.

5.2.1 Experimental design

The experiment was constructed by the author in 2017, and the purpose of it was to elicit focused data on how the antecedent influences the use of epicene pronouns, especially the use of singular *they*. The online experiment was designed to take about 10 to 20 minutes to complete. To make sure that the

answers were as spontaneous as possible, the participants were instructed not to think too much about their responses, but to write what first comes to mind. The experiment was followed by a short background questionnaire, with questions about age, gender, L1 background and time spent in an English-speaking country (see the Appendix to Article 2).

The experiment consisted of three parts: (1) a cloze test (gap-fill task), (2) a sentence-completion task, and (3) a text-completion task. The tasks were designed so that the language in the tasks would represent media discourse that is aimed at the general public. The first task included three text extracts with six gaps in each text. The texts were adaptations of online articles about the use of smartphones, relationship breakups, and mental health nursing. The second task consisted of ten decontextualized, half-finished sentences which the participants were asked to complete. For instance, the task included the following two incomplete sentences: “*My forgetfulness may annoy my family, but I...*” and “*We live in a world where anyone can do exactly what...*”. This task was modelled after a similar task in Abudaljuh (2012). The third task included a text extract from an online opinion piece, where the final sentences had been omitted. The participants were instructed to imagine that they were the author of this text and write an ending to it in three to five sentences.

In total, the experiment included 29 items; of these, 8 items were relevant to the study, that is, they were designed to elicit the use of epicene pronouns. The rest of the items served as distractors so that the participants would not know what the actual purpose of the experiment was. The relevant items included antecedents of the four major types: indefinite pronouns, quantificational NPs, indefinite NPs, and definite NPs. Table 7 shows all relevant antecedents in the order that they occurred in the experimental tasks.

Table 7. Antecedents in the experiment in Study 2

Antecedent	Type	Notional number	Task
<i>the average user</i>	definite NP	singular	1
<i>everybody</i>	indefinite pronoun	plural	1
<i>the mental health nurse</i>	definite NP	singular	1
<i>a prisoner</i>	indefinite NP	singular	2
<i>every employee</i>	quantificational NP	plural	2
<i>a person</i>	indefinite NP	singular	2
<i>anyone</i>	indefinite pronoun	plural	2
<i>someone</i>	indefinite pronoun	singular	3

Two of these items are stereotypically associated with one gender – *the mental health nurse* (female stereotyped) and *a prisoner* (male stereotyped) – while the rest of the items are neutral as regards gender expectancy. The two stereotyped items were chosen in accordance with ratings from previous studies of English nouns: Abudaljuh 2012; Carreiras et al. 1996; Doherty & Conklin 2017; and Kennison & Trofe 2003. These nouns were also shown to be stereotypically male

and female in a study by Misersky et al. (2014), which, in addition to English, examined gender-expectancy of nouns in Czech, French, German, Italian, Norwegian, and Slovak. It was therefore assumed that the two nouns would generally evoke associations of a specific gender, regardless of the participants' L1 background.

5.2.2 Data collection

The experiment was piloted twice at Åbo Akademi University, the only Swedish-speaking university in Finland. The pilot experiments were completed by 79 first-year students who attended an obligatory course in English. After the first pilot test, some tasks and items were modified slightly so that a larger number of relevant answers could be elicited with the experiment.

The final version of the experiment was distributed in the form of an online survey. The experiment was completed in late 2017 and early 2018 by students of English Language and Literature in the Czech Republic, Denmark, Finland, France, Germany, Italy, Russia, Sweden, and Switzerland. All participants were in their first or second year at university. The experiment was targeted at this group of English learners so that the results of the study could be compared to the other substudies focusing on EFL. In total, 338 students participated in the experiment, which elicited 2,205 instances of epicene pronouns. The students were recruited through contact persons at different universities, who posted information about the experiment on online notice boards or sent out information about it on student e-mailing lists. The students completed the experiment in their own spare time.

5.3 Methods of analysis

The analysis of the data followed a relatively similar pattern across the substudies, and it included both quantitative and qualitative techniques. A popular method within learner corpus research is Contrastive Interlanguage Analysis (CIA), a term first introduced by Granger (1996). CIA has traditionally included two types of comparisons: (1) comparing L2 learners and L1 users, and (2) comparing L2 learners with other L2 learners, typically with the L1 background as the differentiating factor. In recent years, researchers have called into question the viability of making comparisons between L1 and L2 English (see e.g. Brutt-Griffler & Samimy 2001). The criticism has mainly focused on the validity of an "L1 speaker norm"; in an era where the number of L2 speakers exceed that of L1 speakers, it is not self-evident that L1 speaker use should be viewed as the target. Granger (2015: 17) has therefore proposed a revised model of CIA, which includes comparisons between L2 varieties and 'reference language varieties', instead of L1 varieties. In addition to inner-circle Englishes, reference language varieties may include outer-circle varieties of various kinds, in line with what has been suggested by advocates of ELF. Nonetheless, learner corpus research tends to have the applied goal of informing language-teaching

professionals about areas of language use that are, or are not, problematic for specific L2 groups. Given that EFL teaching is still largely based on norms set by inner-circle varieties, comparisons between L2 learners and L1 users thus continue to have their place in CIA (see e.g. Granger 2002: 13).

In all substudies of the present dissertation, comparisons were made between different groups of language users (see also Table 4 above). One such grouping was the L1 of the language users, in line with CIA (Granger 2015). Two kinds of comparisons were made based on L1 background: L1 speakers vs L2 speakers (Study 1), and L2 speakers vs L2 speakers (Studies 1-3). As the L1 groups were matched in terms of proficiency, it was assumed that any systematic differences in pronoun use between L1 groups would point to the L1 lingua-cultural background being an influential factor. In the qualitative analysis, the participants' L1s were further categorised into four groups according to how the languages express gender: (1) genderless languages, (2) languages with natural gender, (3) languages with a combination of natural and grammatical gender, and (4) languages with pervasive grammatical gender (see Section 3.1.4). This categorisation is similar to the taxonomy presented in Gygas et al. (2019). The structural features of the L1 were investigated taking into consideration whether the languages have masculine generics and to what degree such use may result in a pervasive male bias in L2 English use.

A second type of categorisation of language users was applied in Study 4, in which group comparisons were based on the two academic fields that the articles were published in: (1) languages and linguistics, and (2) library and information science. In Studies 1-3, it would also have been a possibility to make comparisons between male and female individuals, as information about participant gender was available for these data. Some studies have suggested that there are differences between male and female individuals in their use of gendered language. For instance, Rubin et al. (1994) found that male individuals produced more sexist language than female individuals. However, as no significant differences between male and female participants were found in Studies 1-3, the influence of participant gender was not analysed further in the dissertation.

The type of comparisons that were made remained similar across group categorisations in the four studies. The groups were quantitatively compared in terms of: (1) how frequently they used each pronoun type, (2) with what kinds of antecedents each type of pronoun occurred, and (3) how much individual variation in pronoun use there was in each group. In all substudies, statistical measures were applied to examine differences between the relevant groups. The data were analysed using IBM SPSS Statistics, versions 23 to 25 (see e.g. Brace et al. 2012; Field 2009). Two non-parametric tests were employed: Mann-Whitney U (comparisons between two groups) and Kruskal-Wallis (comparisons between more than two groups). A benefit of using Mann-Whitney U and Kruskal-Wallis is that these statistical tests also take into account inter-individual variation; this is not the case with simple chi-square testing, which has often been employed in corpus research (see e.g. McEnery & Wilson 2001). The basis for choosing non-parametric tests instead of the parametric equivalents t-test and ANOVA was

two-fold: (1) non-parametric tests can be used when the data are not normally distributed, and (2) non-parametric tests are more robust to outliers in the data (see e.g. Field 2009). Based on these criteria, non-parametric tests were judged to be the most suitable options. In all cases, the alpha level of significance was set at $p < .05$.

As described above, the relationship between the epicene pronouns and their co-referential antecedents was analysed in each of the four studies. All antecedents were therefore manually categorised into the four types in SPSS: indefinite pronouns, e.g. *anyone*; quantificational NPs, e.g. *any person*; indefinite NPs, e.g. *a person*; and definite NPs, e.g. *the person*. Each pronoun instance was also manually tagged for type (*he*, *he or she* forms, and *they*), so that pronoun type could be cross-tabulated with antecedent type. To examine the connection between the gender-expectancy of the antecedent and the choice of epicene pronouns, stereotype analysis was conducted in Study 3 and Study 4. The procedures undertaken followed Paterson (2014: 69-74). The gender-expectancy of head nouns in NP antecedents were compared against how speakers of English, both L1 and L2 speakers, have rated them in previous research: Abudaljuh (2012), Carreiras et al. (1996), Doherty & Conklin (2017), and Kennison & Trofe (2003). In these studies, the participants were asked to rate non-gendered human nouns in English according to whether they are male, female, or gender-neutral in terms of gender-expectancy. Where possible, the antecedent nouns in Study 3 and Study 4 were divided in a similar way, and they were then correlated against the use of pronouns with each noun. Not all nouns appearing in the data in Study 3 and 4 were included in the rating lists provided in previous research, and they were therefore not included in the stereotype analysis. If clear near-synonyms existed, these were used in the analysis; an example from Study 3 is the noun *individual*, which did not occur in any of the previous studies. However, the near-synonym *person* was included in previous research and this noun was rated as gender-neutral. Thus, the noun *individual* was also categorised as neutral in terms of gender-expectancy in Study 3. It should be noted that the context in which an antecedent occurs may evoke stereotypes about the gender of the referent even though the antecedent in itself is gender neutral. For instance, the antecedent *a person* might be perceived as stereotypically male if it is used when discussing criminal offence. The immediately preceding textual co-text of the pronouns was therefore analysed for possible stereotypes or triggers of stereotypes.

6 Summaries of the studies

This chapter presents a summary of each of the four articles included in the dissertation. While all articles examine the use of epicene pronouns, they do so from somewhat different perspectives. The first three articles focus on the use of epicene pronouns in English as a foreign language, taking into account possible influence from the learner L1 background. In the fourth and final article, the scope is broadened to lingua franca use of English, where both L1 and L2 uses appear.

6.1 Article 1: Epicene pronouns in intermediate to advanced EFL writing

The first study is a corpus-based analysis of the use of the epicene pronouns *he*, *he or she*, and *they* in L1 and L2 argumentative essays and literature examination papers. Comparisons are made (1) between L1 and L2 writers, and (2) between different groups of L2 writers, based on their L1 background. Where appropriate, statistical tests are employed to detect significant differences between groups. The study also examines what influence the antecedent type has on the choice of epicene pronouns in the various groups.

The data were extracted from two relatively comparable corpora of university student essays written between 1991 and 2004: the ICLE and LOCNESS. The L2 writers come from a total of 13 L1 backgrounds: Bulgarian, Czech, Dutch, Finnish, French, German, Italian, Norwegian, Polish, Spanish, Swedish, Russian, and Turkish. These languages represent three different types of gender systems, that is, genderless languages (Finnish and Turkish); languages with natural gender and grammatical gender (Dutch, Norwegian, and Swedish); and languages with pervasive grammatical gender (Bulgarian, Czech, French, German, Italian, Polish, Spanish, and Russian).

The overall results of the study show that generic *he* was the most common pronoun in the L2 texts (59%, $N = 3,003$), whereas singular *they* was the most common one in the L1 texts (50%, $N = 335$). *He or she* variants were used with a similar frequency in L2 and L1 writing (31% and 26% respectively). Both the L2 and the L1 writers proved to be sensitive to the type of antecedent in their choice of epicene pronouns: the use of generic *he* was most frequent with definite NPs and least frequent with indefinite pronouns, while the use of singular *they* showed the opposite distribution.

Further, differences were detected between the L2 English groups in their uses of generic *he* and singular *they*. The most frequent users of generic *he* were the L1 Russian, Italian, French, and German students (more than 70% of all pronouns in each group). These L1s have highly pervasive grammatical gender systems, with a traditional use of masculine generics. The reason for the frequent use of *he* in these groups may thus, at least partially, be influence from the L1 lingua-cultural background. However, crosslinguistic influence cannot be the

only explanation for variation between the L2 groups as the least frequent users of generic *he* (less than 50%) included speakers of L1s with a variety of gender systems, such as Turkish (genderless), Polish (grammatical gender), and Norwegian (natural and grammatical gender).

6.2 Article 2: Singular *they* in English as a foreign language

The second study is a focused analysis of the use of epicene pronouns in advanced L2 English. In particular, the study aims at investigating how the frequency of singular *they* is affected by (1) the antecedent with which the pronoun co-occurs, and (2) the L1 background of the language learners. The study draws on data from an elicitation experiment, which was completed in 2017-2018 by 338 university students from eight L1 backgrounds: Czech, Danish, Finnish, French, German, Italian, Swedish, and Russian. These L1 backgrounds represent a variety of gender systems, in line with the ones in Article 1. The experiment included antecedents differing in terms of three aspects: definiteness, notional number, and gender expectancy. All three of these aspects relate to the concept of individuation, which describes how salient and figure-like a particular referent is.

The findings of the study reveal that singular *they* was the preferred pronoun overall (59.5%, $N = 2,205$), but the frequency of the pronoun varied both in relation to antecedent type and to L1 group. Singular *they* was most frequent with antecedents including quantifiers (indefinite pronouns and quantificational NPs) and least frequent with definite NPs, both in the overall data and in the different L1 groups. The gender expectancy of the antecedent also proved to be highly relevant to the choice of pronouns: in the overall data, singular *they* was used more than 60% of the time with all antecedents except for the female stereotyped NP *the mental health nurse* (47.5%) and the male stereotyped NP *a prisoner* (28.0%).

Considerable differences in the frequency of *they* were found between the L1 groups. The most frequent users of *they* were the L1 Danish, L1 Swedish, and L1 Finnish students, who used *they* more than 80% of the time. In contrast, the lowest frequencies of *they* were found in the L1 Russian and L1 Italian groups (31.8% and 30.4%, respectively). In these groups, generic *he* was used more often than *they*. A possible explanation for these differences is impact from the L1 background as masculine generics are common in some languages with pervasive grammatical gender, including Russian and Italian.

The results also suggest that there are differences between L1 groups in their perception of singular *they*. In the groups with the lowest frequencies of *they*, the pronoun predominantly occurred with semantically plural antecedents. This finding indicates that the language learners rely on singular *they* as an extended form of plural *they*, which is used when a singular pronoun would cause a mismatch in notional number. In the groups where singular *they* was very frequent, however, the pronoun co-occurred with all kinds of antecedents. It can

thus be assumed that learners in these L1 groups perceive *they* as a singular pronoun in its own right.

6.3 Article 3: Language change in L2 academic writing: The case of epicene pronouns

The third study is a diachronic corpus study, which focuses on the use of the epicene pronouns *he*, *he or she*, and *they* in L2 student academic writing. The data were extracted from the BATMAT corpus (Lindgrén 2016) and from the corpus VESPA-SE (Larsson 2015), both of which comprise student academic papers. All writers are L1 speakers of Swedish who study English Language and Literature at university. The study is divided into two parts: (1) a diachronic analysis of papers produced between the 1970s and the 2010s, using data from BATMAT, and (2) an in-depth analysis of current pronoun use, with data from both BATMAT and VESPA-SE. The study also examines how antecedent type and antecedent stereotyping affect the choice of epicene pronouns.

The results of the study indicate a diachronic shift in the use of epicene pronouns in L2 student academic writing. Whereas generic *he* was the most common pronoun in the BATMAT texts from the 1970s (97%, $N = 33$), it was the least frequent type of pronoun in the 2010s (14.1%, $N = 92$). The frequency of singular *they* showed the opposite pattern: singular *they* was the most common epicene pronoun in the texts from the 2010s (56.5%, $N = 92$) but it merely occurred in one instance before the 1990s. The use of singular *they* was only attested with indefinite pronouns and quantificational NPs in the early decades, but the pronoun later (2000s and 2010s) also co-occurred with the other two antecedent types, definite and indefinite NPs. Most antecedents in the texts were neutral in terms of gender expectancy, and no connection was thus found between the choice of epicene pronouns and gender stereotyping of the antecedent.

As argued in the study, the clear decrease in the use of generic *he* shows that feminist language planning has had an effect also in L2 English. Moreover, the writers are likely to be familiar with gender-fair language use from their L1 background, as issues relating to language and gender have received considerable (media) attention in Swedish-language communities in recent times. Apart from singular *they*, *he or she* variants were frequent alternatives to generic *he* in the data of current use, which is concomitant with what previous research of academic writing has found (see e.g. Parini 2012).

6.4 Article 4: Gendering in open access research articles: The role of epicene pronouns

The fourth study is a corpus-based analysis that examines how the use of epicene pronouns contributes to gendering in open access (OA) journal articles in

English. The study focuses on the epicene pronouns *he*, *he or she*, and *they*, and combines quantitative methods with a context-specific discursive approach. The data are extracted from a corpus of articles published in 40 OA journals in (1) languages and linguistics, and (2) library and information science. The materials also comprise author guidelines from the OA journals included in the study, as well as from a selection of subscription-based journals, for comparison.

The study reveals that there is great variation in the use of epicene pronouns in OA articles. The most common pronoun proved to be singular *they* (46.8%, $N = 1,173$) although *he or she* variants were common (38.5%) as well. The use of generic *he* also occurred with some frequency (14.7%), despite feminist efforts to eradicate generic *he* due to its gender-biased nature. The choice of epicene pronouns was shown to be affected by the type of antecedent in the same way as in the other three studies: singular *they* was most common with quantificational antecedents, and least common with definite NPs. Taken together, these findings demonstrate that the generic individual is gendered (using *he* or *he or she* variants) in more than half of the cases in the OA articles.

As far as the author guidelines are concerned, most journals included in the study do not provide authors with any specific instructions for the use of inclusive language. In the few cases where these journals have guidelines, there is no agreement on how best to avoid gender-biased language use in English. The study also revealed that policy and practice do not necessarily go hand in hand as the use of epicene pronouns in a given journal did not always match the statements given in the journal guidelines.

The study shows that the use of epicene pronouns in OA articles is in a state of instability and flux, and that both language internal (e.g. antecedent type) and language external (e.g. prescriptive grammar) factors may contribute to the variation in use.

7 Results

This chapter presents the central findings of the four substudies and makes comparisons between the different data sets. The findings are also brought into relation with prior research on the use of epicene pronouns. The presentation of the results takes its point of departure in the three main research objectives, as restated below (cf. Section 1.2).

1. What differences, if any, can be seen between L2 English groups with different L1s in terms of their use of epicene pronouns in English writing?
2. How, if at all, do different aspects of the antecedent affect the choice of epicene pronouns in EFL and ELF writing?
3. What kinds of, if any, diachronic changes can be observed in the use of epicene pronouns in L2 English writing?

The chapter is structured according to five central aspects of variation found in the data: (1) variation of pronoun types, (2) variation according to type of antecedent, (3) variation according to research material, (4) variation according to L1 lingua-cultural background, and (5) variation at the individual level.

7.1 Pronoun types across data sets

Table 8 shows the distribution of epicene pronouns in the overall data sets in each of the substudies. The figures in brackets represent percentages of use of each pronoun type in the data sets. For the experimental study, the alternative 'Other' is included in addition to *he*, *he or she* forms and *they* as some students used other pronominal forms than these three pronoun types. The corpus-based studies were limited to examining *he*, *he or she* forms and *they*.

As the table reveals, the distribution of the different types of pronouns vary to a considerable degree between the four data sets. In particular, the ICLE corpus (Study 1) stands out from the other sets of data, exhibiting a markedly higher frequency of *he* (56.0%, $N = 3,003$) and a lower frequency of *they* (15.1%) than the rest of the corpora, as well as the experiment. These findings are indicative of an ongoing diachronic shift in the use of epicene pronouns as the ICLE materials were collected in 1991-2004 while the other materials are all from the 2010s.

Table 8. Pronoun types in the overall data sets

	ICLE	Experiment	BATMAT/ VESPA-SE	COAJA
	Student argumentative essays	Elicited pronoun use	Student academic writing	Published research articles
<i>he</i>	1,681 (56.0)	506 (22.9)	22 (9.9)	173 (14.7)
<i>he or she</i>	867 (28.9)	215 (9.8)	87 (39.2)	451 (38.5)
<i>they</i>	455 (15.1)	1311 (59.5)	113 (50.9)	549 (46.8)
<i>other</i>	-	173 (7.8)	-	-
Total	3,003	2,205	222	1,173

There is also variation in the use of *he or she* forms between the data sets although it is somewhat less striking than for the other types of pronouns. A particularly low frequency of this pronoun type is found in the experimental data of Study 2 (9.8%, $N = 2,205$), which, on the other hand, display the most frequent use of *they*. Some previous studies have excluded *he or she* variants (e.g. Gerner 2000; Laitinen 2007; Paterson 2011, 2014), stating that these types of epicenes were very rarely used in the data sets at hand. This argument is clearly not valid in the present dissertation, as *he or she* forms do, in fact, account for a sizeable proportion of pronoun instances in the corpus materials (see also Section 7.3.2).

7.2 Variation according to antecedent

The results of all four substudies show that the co-referential antecedent has bearing on the language user's choice of epicene pronouns. In particular, two aspects of the antecedent are found to be important: antecedent type and gender stereotyping.

7.2.1 Antecedent type

In all data sets, the antecedents were categorised into the four major types: indefinite pronouns, e.g. *anyone*; quantificational NPs, e.g. *any student*; indefinite NPs, e.g. *a student*; and definite NPs, e.g. *the student*. Table 9 shows the distribution of epicene pronouns according to these antecedent types in the overall data in the three corpus-based studies (Studies 1, 3, and 4). The figures in brackets represent percentages of use of pronoun types in reference to antecedent types.

Table 9. Distribution of pronouns according to type of antecedent in the corpus materials

	Definite NPs	Indefinite NPs	Quant. NPs	Indefinite pronouns
ICLE	<i>N</i> = 955	<i>N</i> = 1,111	<i>N</i> = 187	<i>N</i> = 750
<i>he</i>	601 (62.9)	647 (58.2)	91 (48.7)	342 (45.6)
<i>he or she</i>	290 (30.4)	365 (32.9)	64 (34.2)	148 (19.7)
<i>they</i>	64 (6.7)	99 (8.9)	32 (17.1)	260 (34.7)
BATMAT/ VESPA-SE	<i>N</i> = 87	<i>N</i> = 85	<i>N</i> = 11	<i>N</i> = 39
<i>he</i>	10 (11.5)	11 (12.9)	-	1 (2.6)
<i>he or she</i>	38 (43.7)	47 (55.3)	-	2 (5.1)
<i>they</i>	39 (44.8)	27 (31.8)	11 (100.0)	36 (92.3)
COAJA	<i>N</i> = 574	<i>N</i> = 399	<i>N</i> = 126	<i>N</i> = 74
<i>he</i>	108 (18.8)	49 (12.3)	11 (8.7)	5 (6.7)
<i>he or she</i>	240 (41.8)	169 (42.4)	31 (24.6)	11 (14.9)
<i>they</i>	226 (39.4)	181 (45.3)	84 (66.7)	58 (78.4)

All three data sets exhibit the same general pattern in the distribution of the pronouns *he* and *they*: while generic *he* is more common with definite and indefinite NPs than with antecedents with quantifiers (quantificational NPs and indefinite pronouns), the opposite is true of the frequency of singular *they*. In other words, singular *they* is particularly frequent with antecedents including quantifiers. The use of *he or she* variants does not show as strong a pattern of distribution as generic *he* and singular *they*. Nonetheless, it appears that *he or she* forms behave in a more similar way to *he* than to *they*, in that they are generally more common with antecedents including quantifiers than with definite and indefinite NPs.

The influence of antecedent type was also examined in the experimental data (Study 2). Table 10 shows the distribution of epicene pronouns according to the eight antecedents included in the experiment. The figures represent percentages of use with each antecedent.

Table 10. Distribution of pronouns according to antecedent in the experiment

	<i>N</i>	<i>he</i>	<i>he or she</i>	<i>they</i>	<i>she</i>	Other
<i>the average user</i>	257	23.8	8.9	63.0	-	4.3
<i>everybody</i>	233	9.9	1.7	85.4	-	3.0
<i>the MH nurse</i>	314	3.5	13.7	47.5	27.7	7.6
<i>a prisoner</i>	318	58.5	12.3	28.0	0.3	0.9
<i>every employee</i>	290	23.1	10.0	62.8	0.7	3.4
<i>a person</i>	210	19.1	11.0	63.2	5.7	1.0
<i>anyone</i>	323	5.9	20.1	72.8	0	1.2
<i>someone</i>	260	20.3	13.4	62.5	3.8	-

The two items that elicited the highest frequency of *they* in the data were the indefinite pronouns *everybody* (85.4%, *N* = 233) and *anyone* (72.8%, *N* = 323), a finding which is in line with the results of the corpus-based studies. However, the influence of the antecedent type was less clear for some of the NPs. For instance, the frequency of *they* was very similar with the definite NP *the average user* (63.0%), the indefinite NP *a person* (63.2%) and the quantificational NP *every employee* (62.8%). An explanation for this could be that the experiment only included a limited set of NP antecedents and two of these, *the mental health nurse* and *a prisoner*, were stereotypically gendered (see also Section 7.2.2.).

The diachronic data extracted from the BATMAT corpus (Study 3) show that singular *they* is expanding its territory over time. In the first three decades, the 1970s until the 1990s, the use of singular *they* is only attested with indefinite pronouns and quantificational NPs, antecedents that tend to be notionally plural. To give an example, (11) shows the use of singular *they* with the antecedent *everybody* in the data from the 1990s.

- (11) They also said that if such a grading is used, no spread is obtained, and *everybody* gets a very high and unrealistic mark, compared with what *they* normally get in *their* tests. (BATMAT, Study 3)

In the following two decades, however, singular *they* is also used with the other types of antecedents, that is, definite and indefinite NPs. An example is given in (12), which was extracted from the texts produced in the 2000s. In this example, singular *they* co-occurs with the definite NP *the student*.

- (12) As the tests are administered on university level students, the choice of text should probably try to approximate the type of texts that *the student* faces in *their* studies. (BATMAT, Study 3)

The main results of the substudies are similar to what previous research of both L1 and L2 use has attested (see e.g. Baranowski 2002; Lee 2007; Newman 1992, 1998; Parini 2012; Paterson 2014): the use of singular *they* is particularly

frequent with antecedents including quantifiers, whereas generic *he* behaves in the opposite way, that is, the pronoun is more common with definite and indefinite NPs than with quantificational antecedents.

7.2.2 Gender stereotyping

The influence of gender stereotyping was examined in four data sets: (a) in the experiment (Study 2), (b) in the diachronic data from the BATMAT corpus (Study 3), (c) in the data of contemporary use from BATMAT and VESPA-SE (Study 3), and (d) in the COAJA corpus data (Study 4).

The results of the experiment show that gender expectancy clearly affects language users' pronoun choice when they are prompted with nouns that are highly stereotyped. The stereotyped nouns that were tested in this experiment, *nurse* (female gender-expectancy) and *prisoner* (male gender-expectancy), elicited a much higher frequency of the pronouns *she* (27.7%, $N = 314$) and *he* (58.5%, $N = 318$) respectively than the gender-neutral antecedents in the overall data (see Table 8 above). Example (13) illustrates the use of generic *he* in co-reference with the antecedent *a prisoner* in the sentence-completion task of the experiment. The words within brackets represent the half-finished sentence that the participants were asked to complete.

(13) [If *a prisoner* behaves in a good way] *he* may be out of prison earlier than expected. (Study 2)

The influence of gender expectancy did not prove to be as strong in the corpus data. In the diachronic data from BATMAT, generic *he* was the pronoun that occurred most frequently with male-stereotyped antecedents, such as *president* and *sports coach*. This finding suggests that gender expectancy may have had some impact on the writers' pronoun choice. However, the generally frequent use of generic *he* in the diachronic data (36.8%, $N = 280$) cannot be explained solely in terms of gender stereotyping as most gender-stereotyped antecedents only occurred once and generic *he* was also used with neutral antecedents, such as *person*. In the data of current use, from BATMAT and VESPA-SE, no interaction between gender expectancy of the antecedent and the choice of pronoun was established. In the COAJA (Study 4), in turn, there were no clear indications, either, that gender-expectancy would have affected the writers' use of epicene pronouns. Almost all the antecedents that co-occurred with generic *he* in these corpora also co-occurred with *he or she* variants as well as singular *they*.

In a similar manner to the present work, previous research reveals a somewhat mixed picture of the influence of gender stereotyping. In Abudaljuh's (2012) elicitation test, L1 speakers of Arabic were found to use generic *he* more often with male-stereotyped antecedents than with neutral or stereotypically female antecedents. Similarly, Pauwels' (2001) diachronic study of spoken Australian English revealed that the male-stereotyped nouns *doctor* and *surgeon* co-occurred more frequently with *he* than the neutral nouns *person* and *individual* across the whole time period (1960s to late 1990s). In contrast with

these two studies, the influence of gender expectancy appeared to be very slight in Paterson’s (2014) study of professional British English writing. However, as with the present corpus studies, the majority of the antecedents in Paterson’s (2014) data were categorised as neutral with respect to gender expectancy.

7.3 Variation according to research material

The findings revealed variation in pronoun use that appears to be related to the type of materials. This subsection makes comparisons between student academic writing and elicited data, as well as between student academic writing and professional academic writing.

7.3.1 Student academic writing versus elicited data

Both the experiment (Study 2) and the study of BATMAT and VESPA-SE (Study 3) include participants whose L1 is Swedish. These materials were collected in the 2010s, and all participants are students of English Language and Literature at university. Table 11 shows frequencies of epicene pronouns in the data extracted from these materials.

Table 11. Comparison between L1 speakers of Swedish in the experiment and in BATMAT/VESPA-SE

	Experiment	BATMAT/VESPA-SE
<i>he</i>	4 (3.2)	22 (9.9)
<i>he or she</i>	10 (7.9)	87 (39.2)
<i>they</i>	108 (85.7)	113 (50.9)
<i>other</i>	4 (3.2)	-
Total	126	222

Despite the uniformity of the L1 speaker group in this case, there are notable differences between the two data sets. The disparities lie mainly in the frequency of *he or she* forms and the frequency of singular *they*. In the experimental data, singular *they* is clearly the preferred pronoun overall, being used in as much as 85.7% ($N = 126$) of all cases; the use of *he or she* variants only accounts for 7.9% in these data. In the corpora of academic writing, singular *they* is, again, the most common pronoun (50.9%, $N = 222$), but there is also extensive use of *he or she* variants (39.2%). The frequency of generic *he* remains low in both data sets, that is, 3.2% in the experimental data and 9.9% in the data from BATMAT and VESPA-SE.

A major difference between the two sets of materials is the level of formality of the writing. The student academic writing in BATMAT and VESPA-SE is considerably more formal than the media discourse that is represented in the tasks of the experiment. Moreover, the students will have worked on and edited the texts included in the corpora of academic writing, while the participants in the experiment were specifically instructed not to think so much about their

answers, but to write what first comes to mind. In addition to these differences, the possibility cannot be excluded that the online experimental setting will also have affected the students' pronoun choice.

The variation found in the data corroborates the findings of previous L1 research, which show that both *he or she* variants and singular *they* tend to vary in frequency depending on the level of formality. While *he or she* forms are rarely used in informal language (see e.g. Curzan 2014; Newman 1992), they seem to have a much wider distribution in more formal genres, such as assessed student writing and academic writing (Curzan 2014; LaScotte 2016; Meyers 1990; Mitchell 1994; Parini 2012). The texts in BATMAT/VESPA-SE belong to both of these categories. The frequency of singular *they*, on the other hand, tends to vary in the opposite way: singular *they* has been shown to be especially frequent in informal language (see e.g. Laitinen 2007; Newman 1992; Rubin et al. 1994) while being less common in formal genres (see e.g. Holmes 1998). Moreover, research has shown that L1 users perceive *he or she* variants to be more appropriate than singular *they* in formal language use, most likely owing to the stigma of ungrammaticality that surrounds singular *they* (Bradley et al. 2019; LaScotte 2016).

7.3.2 Student academic writing versus professional academic writing

The materials for the substudies included three corpora of academic writing: BATMAT, VESPA-SE, and the COAJA. While BATMAT and VESPA-SE comprise student academic writing, the COAJA consists of research articles published in peer-reviewed OA journals. The writers are users of ELF and they thus comprise both L1 and L2 speakers of English from across the world. Table 12 shows a comparison of the frequency of epicene pronouns in the student academic writing in BATMAT and VESPA-SE, and the professional academic writing in the COAJA.

Table 12. Comparison between student academic writers and professional academic writers

	BATMAT/VESPA-SE	COAJA
<i>he</i>	22 (9.9)	173 (14.7)
<i>he or she</i>	87 (39.2)	451 (38.5)
<i>they</i>	113 (50.9)	549 (46.8)
Total	222	1,173

As can be seen, the student writers and the professional writers are very similar in their use of epicene pronouns. The most frequently occurring pronoun in both groups is singular *they*, which is used in 50.9% ($N = 222$) of all cases in the student writing and in 46.8% ($N = 1,173$) in the published research articles. However, the use of *he or she* variants is also common in both the student and the professional writing, accounting for 39.2% and 38.5%, respectively. The frequency of generic *he* is low in both corpora, but *he* is slightly less common in the student data (9.9%) than in the professional data (14.7%).

Previous research on academic writing has not produced as consistent results as those of the present study. Compared to Parini's (2012) study of social sciences textbooks, the use of generic *he* was less frequent and the use of *they* more common in both data sets; in Parini's study, generic *he* accounted for as much as 37% of all pronoun instances ($N = 996$), compared to 33% for *he or she* and 30% for singular *they*. However, the textbooks in Parini's corpus were published between 1995 and 2005, making parts of his data considerably older than the data presented here. Furthermore, a much lower frequency of generic *he* was found in a diachronic study by Hegarty & Buechel (2006), who examined the prevalence of generic *he* in articles published in *American Psychological Association* (APA) journals between 1965 and 2004. In these data, only 30 out of 388 articles (7.7%) contained an instance of generic *he*, and all of these instances were found in articles that had been published before 1985. Unfortunately, full comparisons are difficult in this case as Hegarty & Buechel (2006) did not examine which pronouns were used in place of generic *he*. In Pauwels & Winter's (2004) study of the Philippine components of the *International Corpus of English*, student writers were found to be somewhat more prone to using gender-inclusive language than professional writers – a finding that is echoed in the present research. The difference between the two groups could be related to the age distribution of the writers as it can be assumed that the average age of professional writers is considerably higher than that of student writers; indeed, one sign of ongoing language change is that younger generations tend to use later forms whereas older generations use earlier forms (see e.g. Labov 1994: 43-72 on the apparent-time hypothesis).

7.4 Variation according to L1 lingua-cultural background

The influence of the L1 lingua-cultural background was examined in the ICLE corpus (Study 1) and in the experiment (Study 2), both of which include participants from a range of L1 backgrounds. The data show that there are considerable differences in the use of epicene pronouns between L1 groups, which are discussed in this subsection.

7.4.1 Overall findings

Seven L1 groups were included both in the materials selected from the ICLE corpus (data produced 1991-2000) and the experiment (data produced in 2018): Czech, Finnish, French, German, Italian, Russian, and Swedish. In addition to these languages, the experiment also included L1 Danish speakers. Table 13 shows a comparison between these groups in the two data sets. It should be noted that the sets of data may not be fully comparable as the materials are of different kinds: corpus data and data elicited using an online experiment. However, the language used in both materials is similar in terms of style and level of formality.

Table 13. Comparison between L1 groups in the ICLE and in the experiment

	ICLE	Experiment
L1 Czech	<i>N</i> = 223	<i>N</i> = 120
<i>he</i>	148 (66.3)	13 (10.8)
<i>he or she</i>	52 (23.3)	12 (10.0)
<i>they</i>	23 (10.3)	89 (74.2)
<i>other</i>	-	6 (5.0)
L1 Danish		<i>N</i> = 113
<i>he</i>	-	7 (6.3)
<i>he or she</i>	-	4 (3.5)
<i>they</i>	-	99 (87.6)
<i>other</i>	-	3 (2.7)
L1 Finnish	<i>N</i> = 180	<i>N</i> = 470
<i>he</i>	103 (57.2)	21 (4.5)
<i>he or she</i>	58 (32.2)	38 (8.1)
<i>they</i>	19 (10.6)	398 (84.7)
<i>other</i>	-	13 (2.8)
L1 French	<i>N</i> = 165	<i>N</i> = 221
<i>he</i>	125 (75.8)	62 (28.1)
<i>he or she</i>	18 (10.9)	12 (5.4)
<i>they</i>	22 (13.3)	124 (56.1)
<i>other</i>	-	23 (10.4)
L1 German	<i>N</i> = 269	<i>N</i> = 352
<i>he</i>	193 (71.7)	66 (18.8)
<i>he or she</i>	53 (19.7)	24 (6.8)
<i>they</i>	23 (8.6)	240 (68.2)
<i>other</i>	-	22 (6.2)
L1 Italian	<i>N</i> = 319	<i>N</i> = 655
<i>he</i>	245 (76.8)	277 (42.3)
<i>he or she</i>	59 (18.5)	84 (12.8)
<i>they</i>	15 (4.7)	208 (31.8)
<i>other</i>	-	86 (13.1)

Table 13. (Continued)

	ICLE	Experiment
L1 Russian	<i>N</i> = 262	<i>N</i> = 148
<i>he</i>	212 (80.9)	57 (38.5)
<i>he or she</i>	34 (13.0)	30 (20.3)
<i>they</i>	16 (6.1)	45 (30.4)
<i>other</i>	-	16 (10.8)
L1 Swedish	<i>N</i> = 262	<i>N</i> = 126
<i>he</i>	115 (43.9)	4 (3.2)
<i>he or she</i>	105 (40.1)	10 (7.9)
<i>they</i>	42 (16.0)	108 (85.7)
<i>other</i>	-	4 (3.2)

As these figures show, there are notable differences between the older data (ICLE) and the newer data (experiment), as well as between the seven L1 groups that are represented in both studies. Despite the variation, however, the main pattern of diachronic change is very similar in all groups: the use of generic *he* decreases dramatically while there is a large increase in the frequency of singular *they*. The use of *he or she* variants also decreases in all groups apart from the L1 Russian one.

Although the distributions of *he* and *they* change in all groups, the frequencies of these pronouns in one group relative to the others remain fairly stable between the older and the newer data. For instance, the L1 Swedish speakers are the only group of learners that use generic *he* less than half of the time (43.9%, *N* = 262) in the ICLE data, and these learners remain the least frequent users of *he* in the experiment (3.2%, *N* = 126). The opposite is true of the use of singular *they*, which increases from 16.0% to 85.7% in the Swedish group. The L1 Finnish group also displays a low frequency of generic *he* relative to the other groups, with 57.2% (*N* = 180) in the ICLE and 4.5% (*N* = 470) in the experiment. At the other end of the scale are the L1 Russian and L1 Italian groups, being the top users of generic *he* in both the older data (80.9%, *N* = 262 and 76.8%, *N* = 319 respectively) and the newer data (38.5%, *N* = 148 and 42.3%, *N* = 655 respectively). These two groups also display the least frequent use of singular *they* in both studies.

Compared to previous research of EFL, the use of generic *he* was less common and the use of singular *they* more common in all L1 groups in the experimental data, despite the considerable variation between the groups (see Abudalbh 2012; Lee 2007; Pauwels & Winter 2006). The use of *he or she* variants was generally relatively rare in the data, which is similar to what Abudalbh (2012) found in his study of L1 Arabic speakers.

7.4.2 Antecedent type and gender stereotyping

The experimental data (Study 2) show that there are differences between L1 groups in how often they use each pronoun type with different types of antecedents (see also Section 7.2.1). In the Czech, Danish, Finnish, and Swedish L1 groups, no students used generic *he* with the notionally plural indefinite pronouns *everybody* and *anyone*. In contrast, these two antecedents elicited the use of *he* in the remaining four groups, that is the French, German, Italian, and Russian L1 groups. The highest frequency of *he* with this type of antecedent was seen in the L1 Italian group, in which the antecedent *anyone* elicited 44.2% ($N = 104$) use of this pronoun. Example (14) shows an instance from the sentence-completion task in the experiment, where generic *he* is used in co-reference with *anyone*. The bracketed words represent the half-finished sentence that the participants were asked to complete.

(14) [We live in a world where *anyone* can do exactly what] *he* wants without being punished for *his* crimes. (Experiment, Study 2)

The experimental data were also used to examine how gender stereotyping affects the choice of pronouns in the different L1 groups. Because of the small number of instances of pronouns in some cases, the findings need to be interpreted with caution. Although all groups seemed to be affected by stereotyping to some degree, the effect was stronger in some groups. For instance, the gender-neutral definite NP *the average user* did not elicit the use of *she* in any of the groups, whereas the stereotyped definite NP *the mental health nurse* elicited as much as 47.4% ($N = 97$) use in the L1 Italian group and 45.0% ($N = 20$) use in the L1 Russian group. These findings can be compared with the L1 Finnish group and the L1 Danish group, which only displayed 3.0% ($N = 67$) and 13.3% ($N = 15$) use of *she*, respectively, with *the mental health nurse*. Similar comparisons can be made between the gender-neutral indefinite NP *a person* and the male-stereotyped indefinite NP *a prisoner*. In the L1 Italian group, *a person* elicited 37.5% ($N = 56$) use of generic *he*, whereas *a prisoner* elicited 79.4% ($N = 102$) use of the pronoun. These figures can be contrasted with the results of the L1 Swedish group, where the corresponding figures were 0% ($N = 10$) and 25.0% ($N = 16$). In the L1 Swedish group, the NP *a prisoner* was the only antecedent that elicited the use of generic *he*.

7.5 Variation at the individual level

All four studies show that individual language users are not necessarily consistent in their choice of epicene pronouns but may switch from using one type of pronoun to using another type. Such individual variation occurs both when the language user is referencing different types of antecedents, such as *every user* and *a user* in (15), or, though less frequently, when referencing one and the same antecedent, such as *the viewer* in (16).

- (15) Here *every user* must enter into the system through login or registration page for submitting *their* views and ideas. [...] *A user* must provide *his* credential like user ID and password for logging on into the system/page. (COAJA, Study 4)
- (16) In the end, it is up to *the viewer him/herself* to decide what to watch and what to avoid on *their* TV-screens. (ICLE, Study 1)

Examples of these kinds were found in all data sets, including the L1 data from LOCNESS in Study 1. Thus, it appears that mixing of pronouns is not restricted to L2 learner language. In many cases, the mixing of pronouns seems to be related to the type of antecedent. In (15), for instance, the writer opts for *they* with the quantificational NP *every user*, but then uses generic *he* with the indefinite NP *a user*. As was seen in Section 7.2.1, antecedents including quantifiers tend to be co-referential with singular *they* more often than definite and indefinite NPs.

When more than one pronoun type is used with one and the same antecedent, the most common combination in the data sets is to switch from a more individuated pronoun to a less individuated pronoun type. In practice, this means switching either from *he* to *he or she* or *they*, or switching from *he or she* to *they*. Examples (17) to (19) illustrate these kinds of mixing of epicene pronouns.

- (17) [If *a prisoner* behaves in a good way] *he* might get *his or her* sentence reduced. (Experiment, Study 2)
- (18) There is no any other reasonable explanation why *an experienced teacher* would give up *his*, otherwise noble, profession and resort to taking up some far less exciting job so that *they* are able to keep *their* family. (ICLE, Study 1)
- (19) *The prototypical buyer* could be expected to give reasons to why *s/he* should not buy a specific product, and *they* could also be expected to possess a certain degree of power over the seller. (VESPA-SE, Study 3)

Although there seem to exist certain patterns in the mixing of pronouns, it should be noted that all sets of data also include cases where no specific explanation can be given for the use of more than one type of pronoun.

Mixing of pronouns has also been detected in previous research (see e.g. Balhorn 2004; Hekanaho 2020; LaScotte 2016; Meyers 1993). For instance, in her study of L1 public speech and writing, Meyers (1993: 189) states that “*they* often alternates with coordinates of the *he or she* type within the same sentence”. Moreover, Balhorn (2004: 91) gives a very early example of mixing – from Chaucer’s *Canterbury Tales* – where the author uses generic *he* in the same clause as the antecedent, but then switches to singular *they* in an adverbial clause that follows.

7.6 Summary of the results

The results of the substudies show that the distribution of epicene pronouns is subject to extensive variation across the different data sets – variation that relates to both language-internal and language-external factors (for an overview of influential factors in L1 writing, see Paterson 2014).

Owing to second-wave feminist linguist efforts, there has been a shift in how the issue of epicene pronouns is presented in grammars and style guides: while generic *he* used to be the prescribed solution, grammars and style guides now generally discourage use of this pronoun (see Paterson 2014). As the diachronic analyses show, this fact is reflected to some extent in all L1 groups in the EFL data: there is a clear decrease in the use of *he* between the ICLE materials, collected 1991-2004, and the experiment, collected in 2017. The frequency of generic *he* also decreases in the diachronic data of student academic writing, which includes texts from the 1970s until the 2010s. Nonetheless, *he* is still present to varying degrees in all data sets examined, which in itself is not a surprising finding; as Chambers (2013: 312) notes, “linguistic variants that are well entrenched in the language ... tend to linger”.

The move away from generic *he* in prescriptive grammar was not, however, followed by immediate acceptance of singular *they*; in fact, it is only in very recent times that writing manuals and style guides have started to promote the use of singular *they* (see e.g. *Publication Manual of the American Psychological Association* 2019). The lack of official endorsement of *they* is mirrored in the findings of the substudies, which reveal differences between the data sets depending on the level of formality of the writing. This situation is particularly evident when comparing the L1 Swedish students’ academic writing and L1 Swedish students’ performance in the elicitation experiment: while *they* is clearly the preferred epicene pronoun in the experimental data (85.7%, $N = 126$), *he* or *she* variants prove to be very strong contenders in the academic texts (39.2%, $N = 222$).

Another factor that relates to the pronoun choice is the co-referential antecedent. The general tendency that emerges from the data is that both generic *he* and *he* or *she* forms occur more often with definite and indefinite NPs than with antecedents including quantifiers (quantificational NPs and indefinite pronouns), while singular *they* displays the opposite pattern of distribution. These findings concur with what previous research has found in both L1 and L2 English use (see e.g. Abudaljuh 2012; Lee 2007; Newman 1992, 1998; Paterson 2014). The influence of the antecedent can be tied to the notion of individuation, which describes the saliency of the antecedent, that is, how easily discernible it is as an individual entity. As discussed in Section 4.2, the relevant features of individuation in the context of epicene reference are notional singularity/notional plurality and definiteness/indefiniteness, with the feature to the left being more individuated than the feature to the right. Thus, the results show that *he* and *he* or *she* variants are particularly common with highly individuated antecedents whereas *they* is more likely to co-occur with antecedents that are low in individuation. The mechanism behind this seems

fairly straightforward: the more individuated an antecedent is, the more likely it is that the language user will have a detailed mental image of the referent, and the more likely it is that the user will opt for gendered pronouns. In contrast, antecedents with low individuation, such as *anyone*, have fuzzy boundaries and are therefore difficult to conceptualise as specific entities, which makes the language user more prone to choosing *they*, a pronoun which is devoid of gender connotations.

A factor that is closely related to individuation is gender stereotyping. Antecedents that are highly stereotyped evoke either male or female imagery in the language user's mind, and they are thus more salient and figure-like than gender-neutral referents, much like antecedents that are high in individuation. However, while the different aspects of individuation are discrete categories, gender expectancy is arguably more of a spectrum, as illustrated in Figure 1.



Figure 1. Spectrum of gender expectancy in nouns

At the far ends of the spectrum are lexically gendered nouns, that is, nouns that are inherently male or female, such as *woman* and *brother*. Next to the lexically gendered nouns come nouns with a high gender expectancy, like the male-stereotyped noun *prisoner* and the female-stereotyped noun *nurse*. The middle of the spectrum represents nouns that are relatively neutral in terms of gender expectancy, such as *person* and *child*. It can be expected that the majority of personal nouns in English are found somewhere in the middle.

The findings of the experiment revealed that gender expectancy has a clear impact on the choice of epicene pronouns when language users are prompted with stereotyped nouns: antecedents with high gender expectancy are more likely to co-occur with either *he* or *she* than neutral nouns, depending on the gender with which the noun is associated. Although gender stereotyping was not systematically examined in the ICLE data, the comment in example (20) from these data shows that the writer's choice of pronouns has been steered by gender stereotyping (my italics).

(20) If one are supposed to be able to rehabilitate a notorious criminal, *the criminal himself* have to change *his* attitude (I choose to refer to the criminal as male, since most criminals, at least in Norway, are young males with lack of education) and frame of mind. (ICLE, Study 1)

Thus, the writer appeals to real-world likelihood that the referent is male in explaining their choice to use generic *he*.

In contrast with the experimental data, gender stereotyping did not prove to be a major factor in the choice of epicene pronouns in the three corpora of academic writing. One apparent reason for this is that the antecedent nouns in the academic data tended to be relatively neutral as regards gender expectancy. It can be assumed that there are large differences between text genres as to whether or not stereotyped antecedents occur to any large extent. In student academic writing as well as published research articles, referents are perhaps more likely to be of a more generic kind, especially as academic writing tends to strive for objectivity. Nonetheless, it is also highly likely to depend on the academic discipline what types of nouns function as heads in antecedent NPs.

The findings of the dissertation also revealed differences between L1 groups, particularly in the case of generic *he* and singular *they*. In the experimental data, the L1 groups ordered themselves in the following way with respect to the use of generic *he*: Italian (42.3%, $N = 655$), Russian (38.5%, $N = 148$), French (28.1%, $N = 221$), German (18.8%, $N = 352$), Czech (10.8%, $N = 120$), Danish (6.5%, $N = 113$), Finnish (4.5%, $N = 470$), and Swedish (3.2%, $N = 126$). The groups with the lowest frequency of *he*, less than 10%, represent L1s that are genderless (Finnish) or have a combination of natural gender and grammatical gender (Danish and Swedish). The rest of the L1s are languages with pervasive grammatical gender. In the groups with the least frequent use of *he*, singular *they* was the most prevalent epicene pronoun. Based on these findings, there seems to be a connection between the gender system of the L1 and the choice of epicene pronouns in L2 English.

8 Discussion

In this chapter, the central findings of the four studies are dealt with in light of the theoretical framework. Section 8.1 is a synthesis that draws on the findings of all substudies. Section 8.2 considers the implications of the work. The chapter ends with a discussion of the limitations of the studies in Section 8.3.

8.1 The use of epicene pronouns in EFL and ELF writing: A synthesis

The purpose of this dissertation has been to examine the use of epicene pronouns in EFL and ELF writing. In doing so, a variety of potentially influential factors have been considered. Three interrelated variables emerge as particularly relevant to the choice of epicene pronouns: the L1 lingua-cultural background, the semantics of the antecedent, and linguistic prescriptivism.

The findings suggest that speakers of L1s with pervasive gender are generally (1) more likely to opt for generic *he*, and (2) less likely to use singular *they* than speakers of genderless L1s or L1s with a combination of natural and grammatical gender. The importance of the L1 lingua-cultural background may be explained with the framework of thinking for speaking (Slobin 1996, 2003), which holds that using language makes us accustomed to thinking in certain ways because we are required to access specific grammatical forms repeatedly. This situation is not equivalent to a linguistic deterministic point of view, according to which the language user would be cognitively ‘imprisoned’ by their L1 background. For instance, it would be absurd to suggest that a lack of formal expression of gender in the L1 means that speakers of the language would be unable to conceptually distinguish between male and female referents. However, it is conceivable that the L1 lingua-cultural background does influence the degree to which the language user thinks of a human referent as necessarily gendered. Thinking for speaking may be reflected in L2 use in the form of conceptual transfer – “thinking in L1 for speaking in L2” (Pavlenko 2011: 246) – such as transference of the component [GENDER] from the L1 into the L2. In languages with pervasive grammatical gender, all nouns require agreement in terms of gender; while categorisations may be largely arbitrary in the case of inanimates, the choice of agreeing forms for human referents are typically based on biological gender. Owing to this situation, the dichotomy between male and female is deeply ingrained in the structure of the language. L1 speakers of languages with pervasive gender will be accustomed to thinking of human referents as necessarily gendered as gender forms need to be accessed every time a human being is mentioned. In the context of L2 use, English learners may be less prone to opt for singular *they* if they see the world through the ‘gendered glasses’ of the L1. And conversely, it may be easier for L2 learners to incorporate singular *they* – a gender-neutral pronoun – in their mental lexicon if their L1 allows for human referents to be neutral and non-gendered. Another, perhaps even more crucial

point is that there has been a strong tendency for the masculine form to be used generically in gendered languages, and this is still the prevailing norm in some languages (for an overview, see Bußmann & Hellinger 2001, 2002, 2003; Hellinger & Motschenbacher 2015). Thus, this situation may lead to the automatic, albeit incorrect, assumption that the pronoun *he* in English can denote individuals of any gender.

The learner L1 background may also be relevant to how much the semantics of the antecedent affects the choice of pronouns. To some degree, all L1 groups in the present data were influenced by semantic plurality, as evidenced by the fact that singular *they* was most frequent with antecedents that have a plural meaning in all groups. However, relativity effects could explain why gender occasionally overrides semantic plurality in some L1 groups but not in others: in the experiment, a number of L1 German, French, Italian, and Russian students used generic *he* with notionally plural indefinite pronouns, whereas not a single L1 Swedish, Danish, or Finnish student did so. In languages with grammatical gender, not only personal nouns but also indefinite pronouns are gendered and thus control agreement; indefinite pronouns tend to be inherently masculine when used generically, as in German *jeder* ‘each, everyone (m)’ and French *quelqu’un* ‘someone, anyone (m)’. Thus, conceptual transfer may lead to the use of generic *he* even with antecedents that are plural in their meaning. In a similar way, grammatical gender in the L1 may influence the degree to which gender expectancy affects the choice of pronouns. As with notionally plural antecedents, all L1 groups in the experiment were influenced to some degree by stereotyping in their pronoun choice; for example, the male stereotyped antecedent *a prisoner* yielded more frequent use of *he* than any other item in each of the groups. However, in the L1 German, Russian, French, and Italian groups, the influence of gender expectancy was stronger than in the rest of the L1 groups. A plausible reason for this is that the combination of (a) male stereotyping and (b) the use of the masculine as generic in the L1 leads to a double male bias. This finding corroborates previous research which suggests that effects of stereotyping are stronger in gendered languages when there is a match between grammatical gender and the social gender of the referent (see e.g. Irmen & Roßberg 2004).

It is, nonetheless, worth remembering that the expression of gender in language is not necessarily fixed: as Farrelly & Seoane (2012: 392) conclude, “people alter their use of language in response to social change and people influence social change through their use of language”. As a result of ongoing feminist discourses, many language communities have seen concerted efforts at increasing gender equality with the help of language reform. However, the degree to which reform planning has been successfully implemented in practice differs considerably between L1 language communities, owing to diverse factors such as the prevailing gender ideologies in society as well as the (perceived) ease with which gender-fair language use can be achieved (for an overview, see Bußmann & Hellinger 2001, 2002, 2003; Hellinger & Motschenbacher 2015; see also Milles 2013). This situation may explain why there were large variations in the data sets even between the L1 languages with pervasive gender. For instance,

there is a long tradition of feminist activism as well as feminist linguistics in German and French language communities; in German, in particular, language authorities have been successful in adopting and promoting guidelines for the use of gender-fair language (see e.g. Ivanov et al. 2018). In contrast, debates about gender-fair language have not yet resulted in any far-reaching implications and extensive changes in language use in Italian and Russian (see Doleschal & Schmid 2001; Formato 2016). Consequently, L1 speakers of Italian and Russian may be more prone to thinking of masculine generics as the norm for gender-indefinite reference. It is also worth noting that – despite the longstanding push for gender-fair language in German – the L1 German speakers still used generic *he* to a higher degree than the speakers of L1s with less pervasive gender (Danish, Swedish) or no gender at all (Finnish). This finding suggests that the conceptual representation of gender in the L1 can have an impact on thinking for speaking even when efforts have been made to reform language use – as far as is structurally feasible – in line with changes in societal discourses.

For the L2 user of English, the influence of linguistic prescription and language reform may work on two levels. Firstly, L2 users are likely to be affected by the prevailing norms in their L1s, both in terms of their use of and their attitudes to gendered forms in language; such language attitudes also tie in with broader ideologies of gender (see e.g. Sarrasin et al. 2012). Secondly, L2 users will be guided by the norms of ‘Standard English’. As shown in the present data, the relevance of linguistic prescriptivism is particularly evident in the context of both student and professional academic writing, where singular *they* faces competition from *he or she* variants – the “only strictly grammatical way” (Baron 2020: 27) as traditionalists would call the use of such forms. This finding is not unexpected in the sense that academic writing tends to be relatively conservative (see e.g. Turner 2018). It can only be speculated how differently attitudes towards singular *they* would have developed, had it not been for the adamant rejection of the use of this pronoun in 18th and 19th century prescriptive grammar. Already in 1885, for instance, the American author Fred Newton Scott wrote the following (quoted in Baron 2020: 149):

The word they is being used as a pronoun of the common gender every day by millions of people who are not particular about their language, and every other day by several thousands who are particular.

In recent times, it appears that singular *they* has gained more acceptance, even by individuals who are ‘particular about their language’. Given this development, one might expect an increase in the use of singular *they* also in more traditionally formal genres, such as academic discourse.

The variation found in both L1 and L2 use suggests that the system of pronouns is currently in something of a flux in English (for a discussion of language change, see Chambers 2013); as in any case of language change, a certain degree of variability and heterogeneity is to be expected in language use

(see e.g. Weinreich et al. 1968). A pertinent question is how English will continue to evolve in terms of gender-fair reference, not least considering its global spread. One factor that may affect future developments is language planning. Milles (2013: 115f) outlines five main factors that influence the outcome of feminist linguistic reforms: (1) the linguistic features of the suggested reform, e.g. how easily the reform can be implemented and whether language users can understand the purpose of the reform without specific knowledge about grammar; (2) the central actors behind the suggested reform, (3) the political context, (4) the attitudes of official language authorities, and (5) the amount of media coverage that is given to the reform and the linguistic issue at hand. These five factors are visualized in Figure 2. It should be noted that the factors do not follow a specific order, nor are they ranked in any way according to level of importance.

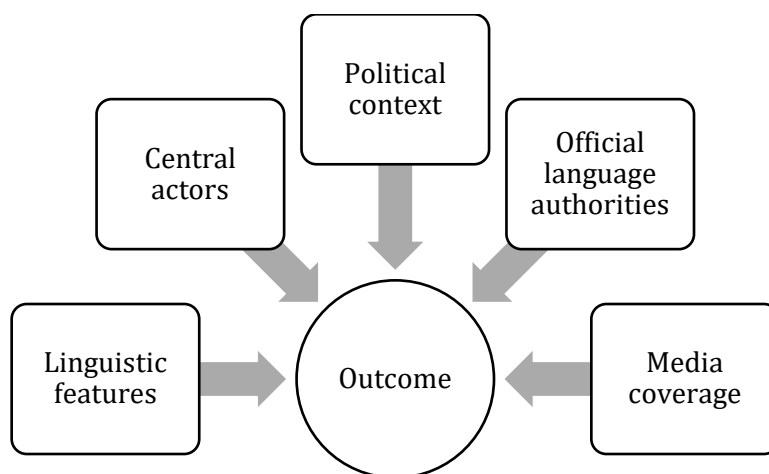


Figure 2. Factors affecting the outcome of feminist language reforms. Adapted from Milles (2013: 115)

All five of these factors contributed in different ways to making the introduction of the new gender-neutral pronoun *hen* in Swedish successful. The implementation of this language reform is a notable development as, typologically speaking, it is rare for a language to have a specific gender-indefinite pronoun if the language makes a distinction between gender in pronouns (see Corbett 1991). In terms of linguistic features, the pronoun *hen* is similar to the existing gendered pronouns (*han* ‘he’ and *hon* ‘she’) and it therefore works well with the syntax of the language. The central actors behind the introduction of *hen* were primarily trans and queer activists as well as gender pedagogues, who did not have a background in linguistics; in other words, the reform can be described as a bottom-up effort. The importance of the political climate is also evident in this context: Sweden has a long and powerful feminist tradition, and the country is rated among the most gender-equal in the world. Thus, the strive for gender-fair language has long been perceived as being in line with “mainstream politics” in Sweden. Moreover, official language authorities in

Sweden were fast at adopting and promoting the use of *hen*, and the new pronoun received extensive media coverage, both within and outside Sweden's borders.

In the case of English, many efforts have been made at introducing a new pronoun; Baron (2020) lists over 200 suggested new epicene pronouns from the 1770s and onwards. However, no neologism has so far spread in a similar fashion to *hen* in Swedish, and it seems rather unlikely that this will happen in the near future (for a discussion, see Baron 2020). Structurally speaking, English and Swedish are very similar with respect to the use of personal pronouns, which means that introducing a new pronoun into English would not be any more difficult than it is in Swedish – at least not from a purely grammatical perspective. However, given the spread of the English language, an enormous number of speakers would have to embrace such a new convention. Another, perhaps more important, factor is that English already has singular *they*, which is a pronoun that: (a) has been around for a long while and is thus familiar to many speakers, and (b) ticks all the right boxes in terms of being inclusive of all genders.

8.2 Implications

The substudies of this dissertation have important theoretical implications for the understanding of the grammatical, cognitive, and social interplay of language users' L1 lingua-cultural background and their L2 English writing as concerns language and gender. They also set a benchmark for future research on the use of personal pronouns in learner and lingua franca varieties of English.

The high prevalence of singular *they* in many English varieties suggests that the use of this pronoun is not going to become obsolete anytime soon; on the contrary, it would seem that singular *they* is currently gaining in popularity. In this context, an empirical question is how singular *they* is stored in English users' mental lexicons. One possibility is that singular *they* is simply a variation or modification of the third-person plural pronoun *they*. This view is supported by Sandford et al. (2007: 373), who describe *they* as a "very tolerant" pronoun that accepts uses with antecedents having differing features, including variations in grammatical number. Another option is that there are two homonymous *they* pronouns – a singular one and a plural one – which are stored as two separate syntactic forms. According to this Homonymy Theory (Whitley 1978; Paterson 2014), singular *they* and plural *they* are similar to singular *you* and plural *you* in terms of how they are represented in the mental lexicon (see Paterson 2014). There are strong indications that the Homonymy Theory is accurate in the case of L1 speakers of English as they use singular *they* in reference to such a wide array of antecedents, including referential ones (Paterson 2014; see also Section 4.1). As far as L2 learners are concerned, the present substudies give a mixed picture: in some L1 groups, most students seem to use singular *they* in a similar fashion to L1 speakers. In other groups, however, *they* mainly occurs with antecedents that are semantically plural, suggesting that these learners have not, in fact, acquired singular *they* as a lexically separate pronoun from plural *they*.

As described above, a possible reason for such differences between L1 groups is that L1s with pervasive grammatical gender do not easily allow for neutralisation of gender. Thus, incorporating the concept of non-gendered personal reference would require some degree of cognitive restructuring for learners with this type of L1 background (see e.g. Pavlenko 2014). A promising path in future research is to examine how singular *they* is acquired and conceptualised by both L1 and L2 speakers of English.

In the early days of second-wave feminism, the debate about language and gender focused mostly on the rights of women and on how female visibility could be increased by using gender-fair forms. In recent years, however, the debate has broadened to include the interests of individuals in the LGBTQ community, in particular the right of gender non-conforming individuals to choose their own pronouns. With this development, singular *they* has expanded its territory to become a singular pronoun of non-binary reference (see e.g. Baron 2020; Bradley et al. 2019). Non-binary antecedents are very highly individuated as they denote specific individuals and are generally named (cf Table 3 above; see also Fraurud 1996). Previous research suggests that L1 users of English tend to be hesitant to use singular *they* with named entities. One reason for this is language users' expectation that the referent "has a gender that should be encoded in the sentence; and furthermore, that this gender is either male or female (or that those are the only genders for which the language provides a means of encoding, and thus one of them must be chosen)" (Bradley et al. 2019: 50). As far as L2 use is concerned, it can be speculated that individuals who do not even use singular *they* with notionally plural antecedents – antecedents that are very low in individuation – are unlikely to do so with the most individuated types of antecedents. Influence from the L1 background and possible relativity effects may be of importance in this context, given that a grammatical masculine-feminine contrast in language puts evident constraints on how individuals may be referenced (see e.g. Motschenbacher 2014). For instance, in comparing English and German, Motschenbacher (2014: 258) concludes that "[a] strictly non-heteronormative use of German is impossible, given that, in many cases, gender-neutral alternatives do not exist (and this is similarly true for many other grammatical gender languages)". However, more research is needed on the use of and attitudes to non-binary *they* in EFL and ELF.

The findings of the dissertation have important practical implications for EFL teaching and the teaching of English for Academic Purposes (EAP). While the L1 may influence L2 use, the structure of the L1 should not be regarded as a straitjacket from which the language user cannot escape. For instance, Whorf ([1940] 2012: 286) believed that the restrictions that the L1 may pose on an individual's thinking can be overcome by consciousness raising in foreign language learning:

If, however, he [the L1 English learner of French] is so fortunate as to have his elementary French taught by a theoretic linguist, he first has the patterns of the English formula explained in such a way that they become semiconscious,

with the result that they lose the binding power over him which custom has given them, though they remain automatic as far as English is concerned. Then he acquires the French patterns without inner opposition, and the time for attaining command of the language is cut to a fraction.

As far as gender is concerned, it is pertinent that learners are made aware of the problematic aspects of using generic *he* in English. Because of its conceptual baggage, this pronoun primarily triggers male-specific imagery, thus excluding a large number of potential referents (for a review, see Henley & Abueg 2003). As a result of this situation, many speakers now regard the use of generic *he* as gender-biased, or even sexist. While a language user may have no intention of expressing themselves in a sexist way by using generic *he*, their audience may still perceive it this way:

Perhaps especially disconcerting for some speakers is that the repercussions of the choice weigh heavily on the connotations an audience may bring to a word. That is, it is not a question only of speaker intention but also of audience reception. (Curzan 2014: 115)

Discussing the issue of gender-bias would be particularly pertinent in L1 groups that display frequent use of generic *he* because of influence from their L1 lingua-cultural background. The results of the substudies strongly suggest that speakers of L1s with pervasive grammatical gender are an important target group in this respect. As indicated by previous research, presenting language users with arguments for gender-fair language can influence them to utilise more gender-fair forms themselves (see e.g. Hansen et al. 2016; Koeser & Sczesny 2014).

Awareness raising is also vital in the context of academic writing. Previous research has suggested that reading texts that are written using gender-fair language can increase individuals' own use of gender-fair forms (Hansen et al. 2016). Given that published research articles serve as models for academic writers across the globe, it would thus be beneficial if journals had a system in place for promoting the use of bias-free language. For instance, author guidelines and submission checklists for editors and reviewers could encourage both authors and gatekeepers to consider this aspect of language use (see also Willis & Jozkowski 2017).

A crucial question is what solution – if any specific – should be recommended in place of generic *he* in teaching as well as in journal guidelines. While linguistic gender equality is a widely shared goal, one cannot ignore that EFL learners and ELF users represent an extensive range of lingua-cultural backgrounds. In many languages with pervasive gender, feminisation strategies tend to be the preferred solution. Feminisation is also possible in English by using combined *he or she* forms. As suggested in previous research as well as in the present dissertation, *he or she* forms are used primarily in formal language. Nonetheless, prescriptive grammar has frequently urged a sparing use of *he or she* variants, on the grounds that they are cumbersome when used repeatedly. This view is

also represented in the EFL corpus data of the dissertation, as exemplified in (21) (my italics).

(21) Now, I think it is time I reveal what I think *an ideal leader* should be like. The age is important because I want *him or her* to have a lot of energy to make changes in society. So *he* can't be too old or too young either because in that case *he* wouldn't have a lot of experience. *He* has to be honest and know that *he* is responsible for *his* own mistakes. [...] Ps. *I want female leaders too but I didn't want to write "he or she" all the time.* (ICLE, Study 1)

In recent times, another issue that has been raised within post-structuralist approaches is that the use of *he or she* variants places unnecessary emphasis on the binary distinction between men and women, as well as excludes individuals who do not identify within this binary (see e.g. Paterson 2020); in Baron's (2020: 7) words, *he or she* forms are simply "too long, too awkward, too binary". Given these drawbacks, it would therefore seem that feminisation does not work as an all-round solution to the issue of epicene pronouns in English.

Another option is to endorse the use of singular *they*, which, as evidenced by previous research, is the most inclusive of the existing third-person singular pronouns (for a review, see Stahlberg & Sczesny 2007). The main caveat is that singular *they* has not been viewed as a singular pronoun proper in prescriptive grammar, but rather as a variation of the plural pronoun, which therefore cannot agree with a grammatically singular antecedent. However, as concluded above, the extensive use of singular *they* in L1 English is an indication that *they* is stored separately from plural *they* in the language user's mental lexicon, thus being a singular pronoun in its own right (see Paterson 2014). It now seems that the persistent scepticism of prescriptive grammarians is slowly subsiding; for instance, singular *they* is the recommended solution in the 7th edition of the *Publication Manual of the American Psychological Association* (APA), which was published in 2019. With this development, journal editors might feel more confident in endorsing the use of *they* in peer reviewing and author guidelines. Also, given that the use of singular *they* is so widespread, it would clearly benefit learners if EFL teaching included information about singular *they* and its possible uses (for similar suggestions, see Brown 2019; Sunderland 1992). In this light, I propose that singular *they* should be included in a modified paradigm of personal pronouns in EFL teaching, as illustrated in Table 14 (see also Bodine 1975 for a related paradigm).

Table 14. Modified paradigm of personal pronouns in English

Person		Number	
		Singular	Plural
1 st		<i>I</i>	<i>we</i>
2 nd		<i>you</i>	<i>you</i>
3 rd	Human	Masculine	<i>he</i>
		Feminine	<i>she</i>
		Unspecified	<i>they</i>
	Non-human	<i>it</i>	<i>they</i>

In this paradigm, *they* occupies two different slots, in a similar way to *you*: once as a third-person plural pronoun, and once as a third-person singular pronoun. The categorisation of *they* as a singular pronoun of unspecified gender is intended to entail both *they* as an epicene pronoun and *they* as a non-binary pronoun.

8.3 Limitations

In any study examining linguistic variation, there will be challenges in disentangling different influential factors from one another. This issue is also a limitation of the present work. For instance, the differences found between the L1 groups could potentially be down to some additional factor apart from the L1 lingua-cultural background. One such factor is the role of EFL instruction, which the substudies of the dissertation were not able to consider to any high degree. EFL instruction involves many potentially influential aspects, ranging from what kinds of gender representations are included in textbooks and other materials to what personal attitudes and ideologies the EFL instructor brings to their teaching. For some learners, the English input they receive at school may constitute a substantial part of their encounters with the language while others may spend hours a day in various online English-speaking settings. From this follows that the role of EFL instruction is very difficult to assess, particularly at advanced levels of proficiency. A related issue is that the status of the English language varies between countries across Europe. In northern Europe, such as Germany and Scandinavia, English has a high status as a foreign language and attitudes towards cultural influences of English are generally positive (see e.g. Lillis and Curry 2006: 6). It may therefore be that learners in these areas are more exposed to English than learners in areas where English does not have as high a profile, such as in southern Europe. Nonetheless, the learners in the substudies focusing on EFL are all university students of English Language and Literature, and it can thus be assumed that these individuals use and are exposed to English in a variety of settings, regardless of their L1 background.

Another limitation of the dissertation is that it focuses on a specific set of L1 backgrounds. While the L1s included show some variation in their gender systems, this small selection of languages cannot in any way capture the

variation that the expression of gender across languages exhibits. With the exception of Finnish and Turkish, all languages represented in the substudies are Indo-European languages. Although they differ from one another to some degree, all these languages share essential similarities in their structures as well as the cultures to which they are tied. However, users of EFL and ELF also include individuals whose L1s have fundamentally different systems of gender. A challenging but promising avenue in future research is thus to further investigate the connection between the L1 gender system and thinking for speaking in L2 English. This question needs to be examined with carefully planned experiments that can identify in what ways the particular gender system influences thinking for speaking in an L2. Future studies also need to include a wider selection of L1 backgrounds, such as L1 speakers of radical pro-drop languages like Chinese and Japanese.

9 Conclusion

This dissertation has added to the body of research on epicene pronouns in English by extending the scope to include non-native varieties. The findings of the substudies reveal that the use of epicene pronouns in EFL and ELF writing is subject to extensive variation and that L2 speakers do not necessarily follow the same usage patterns as L1 speakers. The variation in pronoun use results from several interrelated factors, of which three proved to be particularly important: (1) the language users' L1 lingua-cultural background (formal expression of gender in the L1), (2) the semantics of the antecedent (notional number and gender stereotyping), and (3) linguistic prescriptivism (prescription against the use of singular *they*). Evidence was also found of variation in language users' cognitive models of epicene pronouns, notably in their conceptualisations of singular *they*.

The thesis has important implications for the understanding of the grammatical, cognitive, and social interplay of language users' L1 background and EFL writing. Moreover, it raises the issue of gendered language and linguistic gatekeeping in OA academic ELF publishing as well as proposes a modified paradigm of personal pronouns for EFL teaching.

Svensk sammanfattning

I denna avhandling undersöks användningen av generiska pronomen i engelska som främmande språk (English as a foreign language, EFL) och engelska som lingua franca (English as a lingua franca, ELF). Generiska pronomen är en typ av personliga pronomen som används när korrelatets kön är okänt eller irrelevant. Avhandlingen koncentrerar sig på bruket av de tre vanligaste typerna av dessa pronomen i engelskan: *he*, *he or she*-former, och *they*.

I avhandlingen ingår fyra delstudier, som närmar sig det valda temaområdet från olika infallsvinklar. Materialet består av fem korpusar samt ett experiment. Korpusarna innehåller följande typer av texter: (1) argumenterande uppsatser skrivna av universitetsstudenter, (2) avhandlingar och kursuppsatser skrivna av universitetsstudenter, och (3) vetenskapliga artiklar publicerade i internationella Open Access-tidskrifter. Med hjälp av korpusmaterialet undersöks hur pronomenbruket har förändrats över tid, hur korrelatet påverkar valet av pronomen och hur generiska pronomen används av olika grupper av språkbrukare. Materialet från experimentet syftar i sin tur till att fånga korrelatets inverkan på ett mer djupgående plan. I delstudierna undersöks även hur materialtypen kan påverka användningen av pronomen.

Resultaten visar på en tydlig diakron förändring i pronomenbruket. Det är till exempel tydligt att feministisk språkplanering har haft en effekt, eftersom pronomenet *he* förekommer mycket mer sällan i texterna från 2010-talet än i texterna från 1970- till 2000-talet. Samtidigt som *he* minskar i popularitet blir bruket av pronomenet *they* däremot vanligare. Dessa resultat gäller för alla modersmålsgrupper i delstudierna.

Vad beträffar materialtypen verkar det i synnerhet vara texternas formalitetsnivå som påverkar valet av pronomen. I experimentet, som innehåller relativt informellt språkbruk, är *they* det överlägset vanligaste pronomenet. I de akademiska texterna får *they* däremot konkurrens av *he or she*-former, som används i nästan lika stor utsträckning som *they*.

En tredje faktor som inverkar på pronomenbruket är korrelatet. Pronomenet *they* används i synnerhet med korrelat som står i singular men som har en plural betydelse, såsom de indefinita pronomenen *everyone/everybody* 'alla, var och en'; *he* och *he or she*-former används däremot oftare med korrelat som tydligt syftar på bara en enda individ, såsom *the person* 'personen'. Valet av pronomen tycks också i viss mån påverkas av stereotypa uppfattningar om könsroller: I experimentet används pronomenet *he* oftare med det stereotypiskt manliga korrelatet *prisoner* 'fånge' än med någon annan typ av korrelat. Pronomenet *she* används i sin tur nästan uteslutande med det stereotypiskt kvinnliga korrelatet *nurse* 'sjuksköterska'.

Bruket av generiska pronomen varierar även mellan grupper av språk-användare. Resultaten visar på betydande skillnader mellan universitetsstudenterna som tyder på inverkan från deras modersmål. De studenter som har italienska och ryska som modersmål använder pronomenet *he* oftare än något annat pronomen. Det här pronomenet förekommer däremot

sällan bland studenter med svenska, danska eller finska som modersmål; dessa språkanvändare tycks istället föredra pronomenet *they*.

Variationen som påvisas i de fyra delstudierna tyder på att användningen av generiska pronomen i engelskan befinner sig i ett tillstånd av förändring. En orsak till detta är att bruket av pronomenet *they* under lång tid ratades i preskriptiv grammatik – detta trots att *they* har använts i tredje person singular under flera århundraden. Resultaten pekar också på att modersmålets genusstruktur påverkar användningen av generiska pronomen i EFL. Till exempel verkar språkbrukarna vara mer benägna att använda pronomenet *he* om deras modersmål har grammatiskt genus. Detta resultat kan knytas till språklig relativism, det vill säga teorin om att människors tänkande och världsbild kan påverkas av den grammatiska struktur som deras modersmål har.

References

- Abudalbah, M. 2012. Ideology, gender roles, and pronominal choice: A sociolinguistic analysis of the use of English third person generic pronouns by native speakers of Arabic. (Unpublished doctoral dissertation). University of Kansas, Lawrence.
- Aikhenvald, A. 2016. *How gender shapes the world*. Oxford: Oxford University Press.
- American Psychological Association. 2019. *Publication manual of the American Psychological Association*. Washington, D.C.: American Psychological Association.
- Andersson, E. 2000. 'How many gender categories are there in Swedish?'. In Unterbeck, B. & Rissanen, M. (Eds.) *Gender in Grammar and Cognition. I: Approaches to Gender*. Berlin: de Gruyter: 545–559.
- Baker, P. 2010. *Sociolinguistics and corpus linguistics*. Edinburgh: Edinburgh University Press.
- Balhorn, M. 2004. 'The rise of epicene *they*'. *Journal of English Linguistics* 32(2): 79–104.
- Baranowski, M. 2002. 'Current usage of the epicene pronoun in written English'. *Journal of Sociolinguistics* 6(3): 378–397.
- Baron, D. 1981. 'The epicene pronoun: The word that failed'. *American Speech* 56(2): 83–97.
- Baron, D. 1986. *Grammar and gender*. New Haven: Yale University Press.
- Baron, D. 2020. *What's your pronoun? Beyond he & she*. London: Liveright.
- Berlin, B. & Kay, P. 1969. *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California.
- Biber, D., Conrad, S. & Reppen, R. 1998. *Corpus linguistics. Investigating language structure and use*. Cambridge: Cambridge University Press.
- Bodine, A. 1975. 'Androcentrism in prescriptive grammar: Singular 'they', sex-indefinite 'he', and 'he or she''. *Language in Society* 4(2): 129–146.
- Boroditsky, L., Schmidt, L.A. & Phillips, W. 2003. 'Sex, syntax, and semantics'. In Gentner, D. & Goldin-Meadow, S. (Eds.) *Language in mind: Advances in the study of language and thought*. Cambridge, MA: MIT Press: 61–79.
- Boroditsky, L. 2011. 'How language shapes thought'. *Scientific American* 304(2): 63–65.
- Brace, N., Kemp, R. & Snelgar, R. 2012. *SPSS for psychologists*. New York: Palgrave Macmillan.
- Bradley, E., Schmid, M. & Lombardo, H. 2019. 'Personality, prescriptivism, and pronouns'. *English Today* 35(4): 41–52.
- Bradley, E. 2020. 'The influence of linguistic and social attitudes on grammaticality judgments of singular 'they''. *Language Sciences* 78: 1–11.
- Braun, F., Sczesny, S. & Stahlberg, D. 2005. 'Cognitive effects of masculine generics in German: an overview of empirical findings'. *Communications* 30: 1–21.
- Brown, M.H. 2019. 'The reality and necessity of teaching singular 'they''. *The PanSIG 2018 Journal*: 28–34.
- Brutt-Griffler, J. & Samimy, K. 2001. 'Transcending the nativeness paradigm'. *World Englishes* 20(1): 99–106.
- Bußmann, H. & Hellinger, M. (Eds.). 2001. *Gender across languages: The linguistic representation of women and men*. Vol 1. Amsterdam: John Benjamins.
- Bußmann, H. & Hellinger, M. (Eds.). 2002. *Gender across languages: The linguistic representation of women and men*. Vol 2. Amsterdam: John Benjamins.
- Bußmann, H. & Hellinger, M. (Eds.). 2003. *Gender across languages: The linguistic*

- representation of women and men. Vol 3. Amsterdam: John Benjamins.
- Butler, J. 1990. *Gender trouble: Feminism and the subversion of identity*. London: Routledge.
- Bylund, E. & Jarvis, S. 2011. 'L2 effects on L1 event conceptualization patterns'. *Bilingualism* 14(1): 47–59.
- Carreiras, M., Garnham, A., Oakhill, J. & Cain, K. 1996. 'The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish'. *The Quarterly Journal of Experimental Psychology* 49(3): 639–663.
- Centre for English Corpus Linguistics. 2016. *The Louvain Corpus of Native English Essays (LOCNESS)*. Retrieved from <http://www.learnercorpusassociation.org/resources/tools/locness-corpus> (last accessed July 2020).
- Chambers, J.K. 2013. 'Patterns of variation including change'. In Chambers J.K. & Schilling, N. (Eds.) *The handbook of language variation and change*. Oxford: Blackwell Publishers: 297–324.
- Chen, J-Y. & Su, J-J. 2011. 'Differential sensitivity to the gender of a person by English and Chinese speakers'. *Journal of Psycholinguistic Research* 40(3): 195–203.
- Čmejrková, S. 2003. 'Communicating gender in Czech'. In Bußmann & Hellinger (Eds.): 27–57.
- Coates, J. & Cameron, D. (Eds.). 1988. *Women in their speech communities*. London: Longman.
- Cooper, R. 1984. 'The avoidance of androcentric generics'. *International Journal of the Sociology of Language* 50: 5–20.
- Corbett, C. 1978. 'The agreement hierarchy'. *Journal of Linguistics* 15(2): 203–395.
- Corbett, G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Crystal, D. 2003. *English as a global language*. Cambridge: Cambridge University Press.
- Curzan, A. 2003. *Gender shifts in the history of English*. Cambridge: Cambridge University Press.
- Curzan, A. 2014. *Fixing English: Prescriptivism and language history*. Cambridge: Cambridge University Press.
- Doherty, A. & Conklin, K. 2017. 'How gender-expectancy affects the processing of "them"'. *The Quarterly Journal of Experimental Psychology* 70(4): 718–735.
- Earp, B. 2012. 'The extinction of masculine generics'. *Journal for Communication and Culture* 2(1): 4–19.
- Eckert, P. 2000. *Linguistic variation as social practice: The linguistic construction of identity in Belten High*. Oxford: Blackwell.
- Ekiert, M. 2010. 'Linguistic effects on thinking for writing: The case of articles in L2 English'. In Han, Z. & Cadierno, T. (Eds.) *Linguistic relativity in SLA: Thinking for speaking*. Bristol: Multilingual Matters: 125–153.
- Ellis, R. & Barkhuizen, G. 2005. *Analysing learner language*. Oxford: OUP.
- Engelberg, M. 2016. *Yleispätevä mies: Suomen kielen geneerinen, piilevä ja kieliopillistuva maskuliinisuus*. Helsinki: University of Helsinki.
- Everett, C. 2011. 'Gender, pronouns and thought: The ligature between epicene pronouns and a more neutral gender perception'. *Gender and Language* 5(1): 133–152.
- Everett, C. 2013. *Linguistic relativity. Evidence across languages and cognitive domains*. Berlin: De Gruyter.

- Field, A. 2009. *Discovering statistics using SPSS*. London: SAGE.
- Firth, A. 1996. 'The discursive accomplishment of normality: On 'lingua franca' English and conversation analysis'. *Journal of Pragmatics* 26: 237–259.
- Formato, F. 2016. 'Linguistic markers of sexism in the Italian media: A case study of *ministra* and *ministro*'. *Corpora* 11(3): 371–399.
- Fraurud, K. 1996. 'Cognitive ontology and NP form'. In Fretheim, T. & Gundel, J.K. (Eds.) *Reference and referent accessibility*. Amsterdam: John Benjamins: 65–87.
- Friedrich, P. 1989. 'Language, ideology, and political economy'. *American Anthropologist* 91: 295–312.
- Gastil, J. 1990. 'Generic pronouns and sexist language: The oxymoronic character of masculine generics'. *Sex Roles* 23(11/12): 629–643.
- Gennari, S., Sloman, S., Malt, B. & Fitch, W.T. 2002. 'Motion events in language and cognition'. *Cognition* 83(1): 49–79.
- Gerner, J. 2000. 'Singular and plural anaphors of indefinite personal pronouns in spoken British English'. In Kirk, J. (Ed.) *Corpora galore: Analyses and techniques in describing English: Papers from the nineteenth international conference of English language research on computerised corpora (ICAME 1998)*. Amsterdam: Rodopi: 93–114.
- Gilquin, G. 2007. 'To err is not all. What corpus and elicitation can reveal about the use of collocations by learners'. *Zeitschrift für Anglistik und Amerikanistik* 55(3): 273–291.
- Gilquin, G. & Gries, S. 2009. 'Corpora and experimental methods: A state-of-the-art review'. *Corpus Linguistics and Linguistic Theory* 5(1): 1–26.
- Granger, S. 1996. 'From CA to CIA and back: An integrated contrastive approach to computerized bilingual and learner corpora'. In Aijmer, K., Altenberg, B. & Johansson, M. (Eds.) *Languages in Contrast. Text-based cross-linguistic studies*. Lund: Lund University Press: 37–51.
- Granger, S. 2015. 'Contrastive interlanguage analysis: A reappraisal'. *International Journal of Learner Corpus Research* 1(1): 7–24.
- Granger, S., Dagneaux, E., Meunier, F. & Paquot, M. (Eds.). 2009. *The International Corpus of Learner English. Version 2* (Handbook + CD-ROM). Louvain-la-Neuve: PUL.
- Gumperz, J. & Levinson, S. (Eds.). 1996. *Rethinking linguistic relativity*. Cambridge: Cambridge University Press.
- Gygax, P., Gabriel, U., Sarrasin, O., Oakhill, J. & Garnham, A. 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.
- Gygax, P., Gabriel, U., Lévy, A., Pool, E., Grivel, M. & Pedrazzini, E. 2012. 'The masculine form and its competing interpretations in French: When linking grammatically masculine role names to female referents is difficult'. *Journal of Cognitive Psychology* 24(4): 395–408.
- Gygax, P., Elmiger, D., Zufferey, S., Garnham, A., Sczesny, S., Von Stockhausen, L., Braun, F., & Oakhill, J. 2019. 'A language index of grammatical gender dimensions to study the impact of grammatical gender on the way we perceive women and men'. *Frontiers in Psychology* 10: 1604.
- Halliday, M.A.K. & Hasan, R. 1976. *Cohesion in English*. London: Longman.
- Hamilton, M. 1988. 'Using masculine generics: Does generic *he* increase male bias in the user's imagery?'. *Sex Roles* 19(11/12): 785–799.

- Han, Z. & Cadierno, T. (Eds.). 2010. *Linguistic relativity in SLA: Thinking for speaking*. Bristol: Multilingual Matters.
- Hansen, K., Littwitz, C. & Sczesny, S. 2016. 'The social perception of heroes and murderers: Effects of gender-inclusive language in media reports'. *Frontiers in Psychology* 7: 369.
- Hegarty, P. & Buechel, C. 2006. 'Androcentric reporting of gender differences in APA journals: 1965-2004'. *Review of General Psychology* 10(4): 377-389.
- Hekanaho, L. 2020. *Generic and nonbinary pronouns. Usage, acceptability and attitudes*. Helsinki: Unigrafia.
- Hellinger, M. 1991. 'Feminist linguistics and linguistic relativity'. *Working Papers on Language, Gender and Sexism* 1(1): 25-37.
- Hellinger, M. & Motschenbacher, H. (Eds.). 2015. *Gender across languages: The linguistic representation of women and men*. Vol 4. Amsterdam: John Benjamins.
- Henley, N. & Abueg, J. 2003. 'A review and synthesis of research on comprehension of the masculine as a generic form in English'. *Estudios de Sociolingüística: Linguas, Sociedades e Culturas* 4(2): 427-454.
- Holmes, J. 1998. 'Generic Pronouns in NZ English'. *Kōtare: New Zealand Notes and Queries* 1(1): 32-40.
- Hopper, P. & Thompson, S. 1980. 'Transitivity in grammar and discourse'. *Language* 56(2): 251-299.
- House, J. 1999. 'Misunderstanding in intercultural communication: Interactions in English as a lingua franca and the myth of mutual intelligibility'. In Gnutzmann C. (Ed.) *Teaching and learning English as a global language*. Tübingen: Stauffenburg: 73-89.
- Huddleston, R. & Pullum, G.K. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Humboldt, W. von. 1988 [1836]. *On language: The diversity of human language-structure and its influence on the mental development of mankind*. Bonn: Dümmlers Verlag.
- Ingvardsdóttir, H. & Arnbjörnsdóttir, B. 2013. 'ELF and academic writing: A perspective from the expanding circle'. *Journal of English as a Lingua Franca* 2(1): 123-145.
- Irmen, L. & Roßberg, N. 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.
- Ivanov, C., Lange, M.B. & Tiemeyer, T. 2018. 'Geschlechtergerechte Personenbezeichnungen in deutscher Wissenschaftssprache. Von frühen feministischen Vorschlägen für geschlechtergerechte Sprache zu deren Umsetzung in wissenschaftlichen Abstracts'. *Suvremena Lingvistika* 44(86): 261-290.
- Jakobson, R. 1959. 'On linguistic aspects of translation'. In Brower R. & Fang, A. (Eds.) *On translation*. Harvard: Harvard University Press: 232-239.
- Jarvis, S. 2016. 'Clarifying the scope of conceptual transfer'. *Language Learning* 66(3): 608-635.
- Jarvis, S. & Pavlenko, A. 2008. *Crosslinguistic influence in language and cognition*. New York: Routledge.
- Jenkins, J. 2006. 'Current perspectives on teaching World Englishes and English as a Lingua Franca'. *Tesol Quarterly* 40(1): 157-181.
- Kachru, B. 1985. 'Standards, codification and sociolinguistic realism: The English language in the outer circle'. In Quirk, R. & Widdowson, H.G. (Eds.) *English in the*

- world: *Teaching and learning the language and literatures*. Cambridge: Cambridge University Press: 11–30.
- Kennison, S. & Trofe, J. 2003. 'Comprehending pronouns: A role for word-specific gender stereotype information'. *Journal of Psycholinguistic Research* 32(3): 355–378.
- Kloss, H. 1969. *Research possibilities on group bilingualism: A report*. Quebec: International Center for Research on Bilingualism.
- Koeser, S. & Sczesny, S. 2014. 'Promoting gender-fair language: The impact of arguments on language use, attitudes, and cognitions'. *Journal of Language and Social Psychology* 33(5): 548–560.
- Labov, W. 1994. *Principles of linguistic change. Volume 1: Internal factors*. Oxford: Blackwell.
- Laine, T. & Watson, G. 2014. 'Linguistic sexism in *The Times* – A diachronic study'. *International Journal of English Linguistics* 4(3): 1–22.
- LaScotte, D. 2016. 'Singular they: An empirical study of generic pronoun use'. *American Speech* 91(1): 62–80.
- Laitinen, M. 2007. *Agreement patterns in English: Diachronic corpus studies on common-number pronouns*. Helsinki: Société Néophilologique.
- Lakoff, R. 1975. *Language and woman's place*. New York: Harper and Row.
- Larsson, T. 2015. The Swedish subset of the *Varieties of English for Specific Purposes dAtabase* (VESPA-SE). Compiled at Uppsala University, Sweden. VESPA is administered by Dr. Magali Paquot at the Centre for English Corpus Linguistics at Université catholique de Louvain. <http://www.uclouvain.be/en-cecl-vespa.html>.
- Lee, C. 2017. *Multilingualism online*. New York: Routledge.
- Lee, J. 2007. 'Acceptability of sexist language among young people in Hong Kong'. *Sex Roles* 56: 285–195.
- Leung, C. 2005. 'Convivial communication: recontextualizing communicative competence'. *International Journal of Applied Linguistics* 15(2): 119–144.
- Lewis, G. 2000. *Turkish grammar*. Oxford: Oxford University Press.
- Lillis, T. & Curry, M. J. 2006. 'Professional academic writing by multilingual scholars: Interactions with literacy brokers in the production of English-medium texts'. *Written Communication* 23(1): 3–35.
- Lindgrén, S.-A. 2016. *The BATMAT Corpus*. Turku: The Department of English Language and Literature at Åbo Akademi University.
- Lucy, J.A. 2005. 'Through the window of language: Assessing the influence of language diversity on thought'. *Theoria* 20(54): 299–309.
- MacKay, D.G. & Fulkerson, D.C. 1979. 'On the comprehension and production of pronouns'. *Journal of Verbal Learning and Verbal Behavior* 18: 661–673.
- Madson, L. & Hessling, R.M. 2001. 'Readers' perceptions of four alternatives to masculine generic pronouns'. *The Journal of Social Psychology* 141(1): 156–158.
- Mair, C. 2013. 'The world system of Englishes: Accounting for the transnational importance of mobile and mediated vernaculars'. *English World Wide* 34: 253–278.
- Marcato, G. & Thüne, E.-M. 2002. 'Gender and female visibility in Italian'. In Bußmann & Hellinger (Eds.): 187–218.
- Martyna, W. 1978. 'What does 'he' mean? Use of the generic masculine'. *Journal of Communication* 28: 130–139.

- Martyna, W. 1980. 'Beyond the "he/man" approach: The case for non-sexist language'. *Signs* 5(3): 482–493.
- Matossian, L.A. 1997. *Burglars, babysitters, and persons: A sociolinguistic study of generic pronoun usage in Philadelphia and Minneapolis*. (Unpublished doctoral dissertation). University of Pennsylvania.
- Mauranen, A. 2012. *Exploring ELF: academic English shaped by non-native speakers*. Cambridge: Cambridge University Press.
- Mauranen, A. 2018. 'Second language acquisition, world Englishes, and English as a Lingua Franca (ELF)'. *World Englishes* 37(1): 106–119.
- Mauranen, A., Hynninen, N. & Ranta, E. 2010. 'English as an academic lingua franca: The ELFA Project'. *English for Specific Purposes* 29(3): 183–190.
- McConnell-Ginet, S. 2008. 'Words in the world: How and why meanings can matter'. *Language* 84(3): 497–527.
- McEnery, T. & Wilson, A. 2001. *Corpus linguistics: An introduction*. Edinburgh: Edinburgh University Press.
- Merkel, E., Maass, A. & Frommelt, L. 2012. 'Shielding women against status loss: The masculine form and its alternatives in the Italian language'. *Journal of Language and Social Psychology* 31(3): 311–320.
- Meyers, M. 1990. 'Current generic pronoun usage: An empirical study'. *American Speech* 65(3): 228–237.
- Meyers, M. 1993. 'Forms of *they* with singular noun phrase antecedents: Evidence from current educated English usage'. *Word* 44(2): 181–192.
- Mickan, A., Schiefke, M. & Stefanowitsch, A. 2014. 'Key is a llave is a Schlüssel: A failure to replicate an experiment from Boroditsky et al. 2003'. *Yearbook of the German Cognitive Linguistics Association* 2(1): 39–50.
- Miller, M. & James, L. 2009. 'Is the generic pronoun he still comprehended as excluding women?'. *American Journal of Psychology* 122(4): 483–496.
- Milles, K. 2013. 'En öppning i en sluten ordklass? Den nya användningen av pronomenet *hen*'. *Språk & stil* 23: 107–140.
- Mills, S. 2003. 'Third wave feminist linguistics and the analysis of sexism'. *Discourse Analysis Online*. Available: <https://extra.shu.ac.uk/daol/articles/open/2003/001/mills2003001.html> [accessed July 2020]
- Misersky, J., Gygax, P., Canal, U., Gabriel, A., Garnham, F., Braun, T., Chiarini, N., Englund, A., Hanulikova, A., Öttl, J., Valdrova, L., Von Stockhausen, L. & Sczesny, S. 2014. 'Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak'. *Behavior Research Methods* 46(3): 841–871.
- Mitchell, F. 1994. 'Student writing and pronoun reference: A study of inclusive language in practice'. *Women and Language* 17(2): 2–6.
- Modiano, M. 'International English in the global village'. *English Today* 15(2): 22–28.
- Motschenbacher, H. 2014. 'Grammatical gender as a challenge for language policy: The (im)possibility of non-heteronormative language use in German versus English'. *Language Policy* 13(3): 243–261.
- Motschenbacher, H. 2016. 'A discursive approach to structural gender linguistics: Theoretical and methodological considerations'. *Gender and Language* 10(2): 149–169.
- Moulton, J., Robinson, G. & Elias, C. 1978. 'Sex bias in language use: "Neutral" pronouns that aren't'. *The American Psychologist* 33(11): 1032–1036.

- Newman, M. 1992. 'Pronominal disagreements: The stubborn problem of singular epicene antecedents'. *Language in Society* 21(3): 447–475.
- Newman, M. 1998. 'What can pronouns tell us? A case study of English epicenes'. *Studies in Language* 22(2): 353–389.
- Oakhill, J., Garnham, A. & Reynolds, D. 2005. 'Immediate activation of stereotypical gender information'. *Memory & Cognition* 33(6): 972–983.
- Odlin, T. 2005. 'Crosslinguistic influence and conceptual transfer: What are the concepts?'. *Annual Review of Applied Linguistics* 25(3): 3–25.
- Papafragou, A., Hulbert, J. & Trueswell, J. 2008. 'Does language guide event perception? Evidence from eye movements'. *Cognition* 108(1): 155–184.
- Parini, A. 2012. 'The use of the epicene pronoun in written English: Variation across genres'. In Dontcheva-Navratilova, O. & Povolna, R. (Eds.) *Discourse interpretation*. Newcastle upon Tyne: Cambridge Scholars: 163–178.
- Park, J.S.-Y. & Wee, L. 2012. *Markets of English: Linguistic capital and language policy in a globalizing world*. New York: Routledge.
- Paterson, L.L. 2011. 'Epicene pronouns in UK national newspapers: A diachronic study'. *ICAME Journal* 35: 171–184.
- Paterson, L.L. 2014. *British pronoun use, prescription, and processing. Linguistic and social influences affecting 'they' and 'he'*. London: Palgrave Macmillan.
- Paterson, L.L. 2020. 'Non-sexist language policy and the rise (and fall?) of combined pronouns in British and American written English'. *Journal of English Linguistics* 48(3): 258–281.
- Pauwels, A. 1998. *Women changing language*. London: Longman.
- Pauwels, A. 2001. 'Non-sexist language reform and generic pronouns in Australian English'. *English World-Wide: A Journal of Varieties of English* 22(1): 105–119.
- Pauwels, A. 2011. 'Planning for a global Lingua Franca: Challenges for feminist language planning in English(es) around the world'. *Current Issues in Language Planning* 12(1): 9–19.
- Pauwels, A. & Winter, J. 2004. 'Generic pronouns and gender-inclusive language reform in the English of Singapore and the Philippines'. *Australian Review of Applied Linguistics* 27(2): 50–62.
- Pauwels, A. & Winter, J. 2006. 'Gender inclusivity or 'grammar rules OK'? Linguistic prescriptivism vs linguistic discrimination in the classroom'. *Language and Education* 20(2): 128–140.
- Pavlenko, A. 2011. 'Thinking and speaking in two languages: Overview of the field'. In Pavlenko, A. (Ed.) *Thinking and speaking in two languages*. Bristol: Multilingual Matters: 237–257
- Pavlenko, A. 2014. *The bilingual mind and what it tells us about language and thought*. Cambridge: Cambridge University Press.
- Phillipson, R. 2008. 'Lingua Franca or Lingua Frankensteinia? English in European integration and globalisation'. *World Englishes* 27(2): 250–268.
- Pinker, S. 1989. *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.
- Pinker, S. 1994. *The language instinct: How the mind creates language*. New York: Morrow.
- Prodromou, L. 2007. 'Is ELF a variety of English?'. *English Today* 23(2): 47–53.

- Quirk, R., Greenbaum, S. & Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London, New York: Longman.
- Ramos, S. & Roberson, D. 2011. 'What constrains grammatical gender effects on semantic judgements? Evidence from Portuguese'. *Journal of Cognitive Psychology* 23: 102–111.
- Ringbom, H. 2007. *Cross-linguistic similarity in foreign language learning*. Clevedon: Multilingual Matters.
- Rubin, D., Greene, K. & Schneider, D. 1994. 'Adopting gender-inclusive language reforms. Diachronic and synchronic variation'. *Journal of Language and Social Psychology* 13(2): 91–114.
- Samuel, S., Cole, G. & Eacott, M.J. 2019. 'Grammatical gender and linguistic relativity: A systematic review'. *Psychonomic Bulletin & Review* 26(6): 1767–1786.
- SAOL, Svenska Akademiens ordlista över svenska språket. 2015. 14th edition. Stockholm: Svenska Akademien.
- Sarrasin, O., Gabriel, U. & Gygax, P. 2012. 'Sexism and attitudes toward gender-neutral language'. *Swiss Journal of Psychology* 71(3): 113–124.
- Schafroth, E. 2003. 'Gender in French. Structural properties, incongruences and asymmetries'. In Bußmann & Hellinger (Eds.): 87–139.
- Schmid, U. 1998. 'Zur Bezeichnung weiblicher Personen im Russischen: Eine empirische Pilotstudie'. *Wiener Slawistischer Almanach* 41: 239–262.
- Schneider, E. W. 2007. *Postcolonial English: Varieties around the World*. Cambridge: CUP
- Schneider, E. W. 2014. 'New reflections on the evolutionary dynamics of world Englishes'. *World Englishes* 33(1): 9–32.
- Schriefers, H. & Jescheniak, J. 1999. 'Representation and processing of grammatical gender in language production: A review'. *Journal of Psycholinguistic Research* 28(6): 575–600.
- Scott, M. 2008. *WordSmith Tools Version 5*. Liverpool: Lexical Analysis Software.
- Seidlhofer, B. 2004. 'Research perspectives on teaching English as a lingua franca'. *Annual Review of Applied Linguistics* 24(3): 200–239.
- Sera, M.D., Elieff, C., Forbes, J., Clark Burch, M., Rodríguez, W. & Poulin Dubois, D. 2003. 'When language affects cognition and when it does not: An analysis of grammatical gender and classification'. *Journal of Experimental Psychology* 131(3): 377–397.
- Slobin, D. 1996. 'From "thought and language" to "thinking for speaking"'. In Gumperz, J. & Levinson, S. (Eds.) *Rethinking linguistic relativity*. Cambridge: Cambridge University Press: 70–96.
- Slobin, D. 2003. 'Language and thought online: Cognitive consequences of linguistic relativity'. In Gentner, D. & Goldin-Meadow, S. (Eds.) *Language in mind: Advances in the study of language and thought*. Cambridge, MA: MIT Press: 157–191.
- Spender, D. 1980. *Man made language*. London: Routledge
- Stahlberg, D., Braun, F., Irmen, L. & Sczesny, S. 2007. In Fiedler, K. (Ed.) *Social communication*. Hove: Psychology press: 163–187.
- Strahan, T. 2008. "'They' in Australian English: Non-gender-specific or specifically non-gendered?". *Australian Journal of Linguistics* 28(1): 17–29.
- Stringer, J. L. & Hopper, R. 1998. 'Generic *he* in conversation?'. *Quarterly Journal of Speech* 84: 209–221.
- Sunderland, J. 1992. 'Gender in the EFL classroom'. *ELT Journal* 46(1): 81–91

- Swan, M. 2012. 'ELF and EFL: Are they really different?'. *Journal of English as a Lingua Franca* (1-2): 379–389.
- Tannen, D. 1991. *You just don't understand: Women and men in conversation*. London: Virago.
- Tardy, C. 2004. 'The role of English in scientific communication: Lingua franca or tyrannosaurus rex?'. *Journal of English for Academic Purposes* 3(3): 247–269.
- Timberlake, A. 1977. 'Reanalysis and actualization in syntactic change'. In Li, C. (Ed.) *Mechanisms of Syntactic Change*. Austin, London: University of Texas Press: 141–177.
- Tognini-Bonelli, E. 2001. *Corpus linguistics at work*. Amsterdam: John Benjamins.
- Turner, J. 2018. *On writtenness. The cultural politics of academic writing*. London: Bloomsbury.
- Vigliocco, G., Vinson, D., Paganelli, F. & Dworzynski, K. 2005. 'Grammatical gender effects on cognition: Implications for language learning and language use'. *Journal of Experimental Psychology* 134(4): 501–520.
- Weidmann, U. 1984. 'Anaphoric they for singular expressions'. In Watts, R. & Weidmann, U. (Eds.) *Modes of interpretation: Essays presented to Ernst Leisi on the occasion of his 60th birthday*. Tübingen: Narr: 59–68.
- Weinreich, U., Labov, W. & Herzog, M.I. 1968. 'Empirical foundations for a theory of language change'. In Lehmann, W.P. & Malkiel, Y. (Eds.) *Directions for historical linguistics: A symposium*. Austin: University of Texas Press: 95–188.
- Whorf, B.L. 2012. *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. J. Carroll (Ed.). Cambridge, MA: MIT Press.
- Willis, M. & Jozkowski, K.N. 2017. 'Ladies first? Not so fast: Linguistic sexism in peer-reviewed research'. *The Journal of Sex Research* 55(2): 137–145.
- Wilson, E. & Ng, S.H. 1988. 'Sex bias in visual images evoked by generics: A New Zealand study'. *Sex Roles* 18(3–4): 159–168.
- Wolff, P. & Holmes, K. 2011. 'Linguistic relativity'. *WIREs Cognitive Science* 2(3): 253–265.

Charlotte Stormbom

Gendered language in flux: The use of epicene pronouns in EFL and ELF writing

This work adds to the body of research on epicene pronouns in English by examining the use of these pronouns in English as a foreign language (EFL) and English as a Lingua Franca (ELF). The analyses focus on the pronouns *he*, *he or she*, and *they*, taking into consideration various factors that may influence pronoun use. Three interrelated factors prove to be of particular importance: the language users' first-language (L1) background, the semantics of the antecedent, and linguistic prescriptivism. Evidence is found of variation in language users' cognitive models of epicene pronouns, notably in their conceptualisations of *they*.

The dissertation has important implications for the understanding of the grammatical, cognitive, and social interplay of language users' L1 background and EFL writing. It raises the issue of gendered language and linguistic gatekeeping in academic ELF publishing as well as proposes a modified paradigm of personal pronouns for EFL teaching.

ISBN 978-952-12-4026-3