

HANNA POOK

Pronoun use and variation
in Estonian dialects:
kes ‘who’, *mis* ‘what’
and *keegi* ‘someone’



HANNA POOK

Pronoun use and variation in Estonian dialects:
kes 'who', *mis* 'what' and *keegi* 'someone'



UNIVERSITY OF TARTU
Press

Institute of Estonian and General Linguistics, Faculty of Arts and Humanities,
University of Tartu

Dissertation accepted for the defence of the degree of Doctor of Philosophy on
6th June, 2023 by the Committee of the Institute of Estonian and General
Linguistics, Faculty of Arts and Humanities, University of Tartu.

Supervisor: Professor Liina Lindström, University of Tartu

Opponent: Professor Marja-Liisa Helasvuo, University of Turku

Commencement: August 24, 2023 at 13:00 at Ülikooli 18-204, Tartu

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



European Union
European Regional
Development Fund



Investing
in your future

ISSN 1406-1325 (print)
ISBN 978-9916-27-259-6 (print)
ISSN 2806-2450 (pdf)
ISBN 978-9916-27-260-2 (pdf)

Copyright: Hanna Pook, 2023

University of Tartu Press
www.tyk.ee

ACKNOWLEDGEMENTS

When in the autumn of 2016 I went to ask my then-not-yet supervisor Liina Lindström for advice regarding what I should write my Master's thesis on, I never imagined that the idea of certain pronouns referring to an animacy opposite the one to which they were "supposed to" refer would intrigue me so much that seven years later I would be completing a PhD dissertation on this very same topic. I had always been interested in Estonian dialects, but her suggestion truly opened a new world of pronouns, syntax and variation analysis for me. Now, at the end of my studies, I owe thanks to many people who have made the completion of this dissertation possible.

First and foremost, I do really want to thank my supervisor Liina. I think without her firm guidance I might have never reached the point of writing these words. She managed a perfect balance of encouragement and strictness that always kept me down to earth yet still slowly striving towards my end goal. Thank you so much for all your suggestions, comments, corrections, help and support over these past five years (and before), all of which have only improved my dissertation.

I am very grateful to my internal reviewer Karl Pajusalu for giving me valuable feedback and for noticing even the tiniest of details. I want to thank Uldis Balodis for proofreading the overview article of this dissertation, and Joshua Wilbur and Djuddah Leijen for proofreading the published articles of this dissertation. All of the remaining mistakes are my own. A special thanks goes to Maarja-Liisa Pilvik, who over the years has always been available to answer my many questions on conducting different statistical analyses or illustrating geospatial data. A big thank you goes to the reviewers of this dissertation, Marja-Liisa Helasvuo and Riho Grünthal, for their invaluable comments and suggestions. In addition, this dissertation has been supported by the Centre of Excellence in Estonian Studies (European Union, European Regional Development Fund).

I want to wholeheartedly thank all my friends at Tartu Kassikaitse. These past few years have been the busiest and most stressful years of my entire life and I do not know how I would have done it without your excellent cat-loving company. You showed me that life is about more than just my dissertation, but you also never stopped encouraging me just to keep writing. Another big thank you goes to Enel – our discussions about the weirdness of Estonian and Estonian dialects (among other topics) have always been my favourite and one day when we are old we will get to writing all those articles together.

I am enormously grateful to my family, my mother Ghita, my father Andrus and my sister Kaija, who have always supported me, rooted for me, pushed me, comforted me, taken well-deserved breaks with me and not once stopped believing in me. Finally, there are no words to describe how grateful I am to my partner Henry. During my studies he has always been there to listen to me celebrate and cry, rant and complain (often on topics he didn't understand a word of) even while he was in the midst of the most intense moments in Call of Duty or League of Legends. You have been my rock through everything.

CONTENTS

| | |
|---|-----|
| ACKNOWLEDGEMENTS | 5 |
| LIST OF PUBLICATIONS..... | 7 |
| ABBREVIATIONS..... | 8 |
| 1. INTRODUCTION | 9 |
| 2. PRONOUNS..... | 13 |
| 2.1. Estonian pronouns..... | 13 |
| 2.2. The pronouns <i>kes</i> , <i>mis</i> and <i>keegi</i> from a historical perspective | 16 |
| 2.3. The dialect forms of <i>kes</i> , <i>mis</i> and <i>keegi</i> | 17 |
| 2.4. Animacy in language and the animacy reference of <i>kes</i> , <i>mis</i> and <i>keegi</i> | 26 |
| 2.5. The grammaticalisation of <i>kes</i> , <i>mis</i> and <i>keegi</i> | 30 |
| 3. DIALECTOLOGY AND SYNTACTIC VARIATION | 33 |
| 3.1. Dialectology..... | 33 |
| 3.1.1. Corpus-based dialectology..... | 34 |
| 3.1.2. Dialect syntax | 36 |
| 3.1.3. Dialectology in Estonia..... | 37 |
| 3.1.4. Dialect syntax in Estonia | 40 |
| 3.2. Variation | 42 |
| 4. DATA AND METHODS | 45 |
| 4.1. Data..... | 45 |
| 4.2. Methods | 48 |
| 4.2.1. Conditional inference trees and random forests..... | 49 |
| 4.2.2. Binomial mixed-effects logistic regression | 50 |
| 4.2.3. Multiple correspondence analysis..... | 50 |
| 5. RESULTS AND DISCUSSION | 52 |
| 5.1. Overview of the publications | 52 |
| 5.2. The animacy reference of <i>mis</i> | 55 |
| 5.3. Results from the combined dataset | 60 |
| 6. CONCLUSION..... | 69 |
| SUMMARY IN ESTONIAN | 73 |
| REFERENCES | 81 |
| PUBLICATIONS | 99 |
| CURRICULUM VITAE | 222 |
| ELULOOKIRJELDUS..... | 223 |

LIST OF PUBLICATIONS

- P1. Pook, Hanna (2019). The pronoun *kes* ‘who’ and its referent’s animacy in Estonian dialects. *SKY Journal of Linguistics*, 32, 105–144.
- P2. Pook, Hanna; Lindström, Liina (2022). The use of the indefinite pronoun *keegi* ‘someone’ in Estonian dialects. *Nordic Journal of Linguistics*, 1–32. <https://doi.org/10.1017/S0332586522000221>.
- P3. Pook, Hanna (2021). Object case variation of the pronoun *mis* ‘what’ in spontaneous spoken Estonian and Estonian dialects. *ESUKA-JEFUL*, 12 (1), 259–301. <https://doi.org/10.12697/jeful.2021.12.1.07>.

Contribution to the publications

- P1. Hanna Pook was the sole author of the article.
- P2. Hanna Pook initiated the research. She devised the data coding schema and coded the data. She conducted all the statistical analyses and interpreted the results. As the first author of the study, she wrote the main text of the article and acted as the corresponding author in the publication process.
- P3. Hanna Pook was the sole author of the article.

ABBREVIATIONS

Abbreviations of glosses

| | |
|---------|-----------------------------|
| 1, 2, 3 | first, second, third person |
| ADE | adessive |
| CNG | connegative |
| COM | comitative |
| GEN | genitive |
| ILL | illative |
| INF | infinitive |
| NOM | nominative |
| PL | plural |
| PRT | partitive |
| PST | past tense |
| PTCL | participle |
| SG | singular |
| TRL | translative |

Abbreviations of subdialects

| | |
|-----|-------------|
| Amb | Ambla |
| Jõe | Jõelähtme |
| Jõh | Jõhvi |
| Jäm | Jämaja |
| Kod | Kodavere |
| Krj | Karja |
| Krl | Karula |
| Kul | Kullamaa |
| Lüg | Lüganuse |
| Mar | Martna |
| Pal | Palamuse |
| Pst | Paistu |
| Rid | Ridala |
| SJn | Suure-Jaani |
| Tor | Tori |
| Tõs | Tõstamaa |

1. INTRODUCTION

Variation is an inherent part of language. All languages can vary among speakers, registers, situations, locations, demographic attributes, etc., and this variation can be found on all linguistic levels. Simply put, variation in language exists everywhere. A popular field in variationist linguistics is the study of variation in dialects. It is well known that certain lexical and phonological phenomena are expressed in a variety of ways in geographically different areas: a person from northern Estonia might say *koer* ‘dog’, while a southern Estonian might say *peni* ‘dog’; a dialect speaker from Saaremaa might use *õ* in words where most of Estonia uses *õ*. These kinds of features are easily noticed and recognisable even to non-linguists.

Syntactic variation, on the other hand, is not as easy to spot or even to define. The line between spoken language variation and dialect variation is thin and sometimes it might be difficult to associate certain syntactic phenomena with a dialect background at all. Some syntactic constructions that go against the norms of current Standard Estonian might actually be used so systematically in certain dialects that they have become characteristic of that dialect. These phenomena are often probabilistic, not categorical in nature, and their variation patterns can only be identified by analysing frequency data.

This dissertation is a study of dialectology, the aim of which is to analyse two instances of such syntactic variation in Estonian dialects. Three pronouns are at the central focus of this dissertation: the interrogative-relative pronouns *kes* ‘who’ and *mis* ‘what’, and the indefinite pronoun *keegi* ‘someone, no one’. Morphosyntactically, these pronouns can vary in a multitude of ways in spoken language. Here I have approached their variation from two perspectives: the variation in their selectivity between animate and inanimate referents and the variation in case marking in partial object position.

Firstly, the interrogative-relative pronouns *mis* and *kes* and the indefinite pronouns *keegi* and *miski* ‘something, nothing’ are differentiated by what they can refer to in Standard Estonian: *kes* and *keegi* are only used to refer to animate entities, while *mis* and *miski* are reserved for inanimate entities (Veski 1958: 31; Erelt 2017: 743). However, in some Estonian dialects, the animacy distinction of these pronouns is not as strict and *kes*, *mis* and *keegi* (although not *miski*) can also refer to the entities of the opposite animacy (see examples 1–3).

- (1) Western (Tor)¹
- | | | | | |
|----------------|-------------------|-----------------|----------------|--------------|
| <i>minu</i> | <i>ämmal</i> | <i>ollid</i> | <i>seoksed</i> | <i>rohud</i> |
| I:GEN | mother-in-law:ADE | be:PST:3PL | this kind:PL | medicine:PL |
| <i>kellega</i> | <i>tema</i> | <i>parandas</i> | <i>oma</i> | <i>kopsu</i> |
| who:COM | she | fix:PST:3SG | own | lung:PRT |
- ‘My mother-in-law had this medicine with which she healed her lungs.’

¹ All the examples in this and the following sections come from the Corpus of Estonian Dialects. Each example is preceded by the dialect and subdialect in parentheses. An overview of the corpus is given in Section 4.1.

- (2) Insular (Krlj)
ma oli sōukse inimene mis tahtsi tantsida
 I be:PST:1SG that_kind human what want:PST:3SG dance:INF
 ‘I was the kind of person who wanted to dance.’
- (3) Mulgi (Pst)
nüüd ei ole kedügi seast asja mitte
 now not be:CNG someone:PRT that_kind:PRT thing:PRT not
 ‘Now there is nothing like that.’

Secondly, both Standard Estonian and Estonian dialects exhibit a phenomenon called differential object marking, which means that the distinction between total and partial objects is made on the basis of certain semantic and syntactic features (Aissen 2003; Witzlack-Makarevich & Seržant 2018). In order to use a total object, the situation has to be perfective, the referent quantitatively bounded and the polarity of the clause affirmative. If any of these conditions are not met, a partial object is used. Total objects occur in nominative or genitive, partial objects in partitive (Erelt et al. 1993b: 51–52; Metslang 2017: 264–267; see also Kont 1963; Ogren 2015). Nonetheless, there are a few exceptions to object case variation (see Metslang 2017: 272–273) and the interrogative-relative pronoun *mis* is one of these: in a position where a partial object is expected to occur, partitive *mida* can be replaced with nominative *mis* (see examples 4–5). This variation occurs in written, spoken, dialectal and even old literary Estonian. Interestingly, the pronoun *kes* does not allow such object case variation.

- (4) Võru (Krl)
tuud ei võiq kõnõlda midä ei tiijä
 that:PRT not can:CNG speak:INF what:PRT not know:CNG
 ‘One cannot speak of what one does not know.’
- (5) Eastern (Kod)
mis ma tiin tämäga
 what:NOM I do:1SG s/he:COM
 ‘What will I do with them?’

In this dissertation, Standard Estonian is defined as the official register of Estonian, the norms of which have been constructed based on language planning and prescriptive grammars. However, since the receding of the dialects during the 20th and 21st century, this standard language has increasingly affected other registers of Estonian, meaning that contemporary spoken language is more or less uniform and, in many aspects, very similar to Standard Estonian. For that reason, the norms of the standard language can, in the case of variation in animacy reference, be extended also to contemporary spoken and written language (standardised or not), as this variation is not prevalent in modern speech.

The syntactic variations in selectivity between animate and inanimate referents and in case marking in partial object position are the main objects of research in this dissertation. In earlier literature, linguists have briefly mentioned the possibility of both variations in either Estonian, Estonian dialects or other cognate languages. For the animacy reference of *kes*, *mis* and *keegi*, see, e.g., Alvre (1977: 19, 23, 1987a: 32), Kask and Palmeos (1985: 59), Lonn and Niit (2002: 55), Must and Univere (2002: 271), Tanning (2004: 112) and Viikberg (2020: 174), among others. For the partial object case variation of both the pronoun *mis* and other nouns, see, e.g., Kont (1963: 115), Juhkam (1983: 122–123), Alvre (1986: 7), Must (1987: 287), Viitso and Ernštreits (2012), Metslang (2017: 273), and Winkler and Pajusalu (2018: 96). None of these studies, however, have elaborated on or analysed this use of the pronouns.

Moreover, there have been few studies on the interrogative-relative and indefinite pronouns overall, as they have been often overshadowed by the much more researched personal and demonstrative pronouns. For example, the loss of distinction in animacy that has happened to *kes*, *mis* and *keegi* is not an isolated case, but has occurred also for the personal pronoun *tema* ‘s/he’ and the demonstrative pronoun *see* ‘it’ (see more in Section 2.4), a phenomenon on which a number of studies and articles have been published. This dissertation, therefore, is an important contribution to the study of syntactic properties of the interrogative-relative and indefinite pronouns, broadening the understanding of topics that have not been comprehensively analysed before.

In addition, this dissertation contributes to the field of dialect syntax and corpus-based dialectology, both of which have gained more interest and research possibilities in Estonia only in the last couple of decades. Dialectology has, for a long time, been a field of qualitative or categorical research, but as syntactic variation in dialects is better described as a continuum (rather than categories) of variants, using complex statistical methods in order to obtain graded frequency information allows us to represent the reality of the language in a much better way.

The goal of this dissertation is to study the syntactic variation of *kes*, *mis* and *keegi* in Estonian dialects in order to shed light on their syntactic behaviour and variations from a previously undescribed perspective, as in certain regions or registers these variations can occur very systematically. The dissertation aims to answer the following six research questions:

- R1. Which Estonian dialects and subdialects typically permit reference to inanimate entities with *kes* and *keegi*, and to animate entities with *mis*?
- R2. Which of the studied variables (clause type, case, polarity, function, position) affect the animacy reference of *kes*, *mis* and *keegi*? Under what conditions does reference to the opposite animacy occur?
- R3. How do the geographic and morphosyntactic variables that affect the animacy reference of *kes*, *mis* and *keegi* compare across these pronouns?

- R4. Which of the studied variables (clause type, polarity, tense, use of impersonal voice, verb type, length of the following word) correlate with the speakers' choice to use either nominative *mis* or partitive *mida* in partial object position in Estonian dialects and contemporary speech? Under what conditions does this object case variation happen?
- R5. How do archaic dialects and contemporary speech differ in terms of the variation of *mis* and *mida* in partial object position?
- R6. What are the possible motivations or reasons behind the variation of animacy reference and object case for the pronouns *kes*, *mis* and *keegi*?

This dissertation is comprised of three articles and an overview article. The first and second article (P1 and P2) are concerned with the animacy reference of the interrogative-relative pronoun *kes* and the indefinite pronoun *keegi* in Estonian dialects, answering R1, R2, R3 and R6. The third study (P3) focuses on the object case variation of the pronoun *mis*, both in the dialects and in contemporary speech, adding a comparative angle to the dissertation. P3 answers R4, R5 and R6.

The overview article consists of six chapters, the first of which is this introductory chapter. In Chapter 2, I address the research object of this dissertation: the pronouns. Section 2.1 describes the Estonian pronoun system and gives an overview of previous research. Sections 2.2, 2.3, 2.4 and 2.5 focus specifically on the pronouns of interest in this dissertation – *kes*, *mis* and *keegi* – and discuss their historical development, possible dialect forms, their animacy reference and their grammaticalisation, respectively.

In Chapter 3, I lay out the theoretical approaches for the dissertation. This dissertation is a corpus-based dialectology study, so in Section 3.1, I introduce the history and framework of (corpus-based) dialectology and dialect syntax, both worldwide and in Estonia. As the dissertation examines the syntactic variation in the use of the pronouns *kes*, *mis* and *keegi*, Section 3.2 describes the history of variation studies and their development into modern corpus-based variationist linguistics.

Chapter 4 presents the data and methods. In Section 4.1, I give an overview of the Corpus of Estonian Dialects, which is the basis for analysis in all the publications. Section 4.2 describes the various statistical methods used both in the publications and in this overview article.

Chapter 5 contains the results and discussion of the dissertation. Section 5.1 introduces the three published articles and their results. In Section 5.2, I analyse the animacy reference of the interrogative-relative pronoun *mis*, as this variation is not addressed in the published articles. In Section 5.3, I combined the datasets of all three pronouns to conduct a comprehensive statistical analysis of the variation in animacy reference in all the Estonian dialects. Sections 5.2 and 5.3 also address R1, R2, R3 and R6.

Finally, Chapter 6 sums up the major results and most important outcomes of the dissertation.

2. PRONOUNS

This dissertation deals with the syntactic variation in the use of three pronouns: *kes*, *mis* and *keegi*. In this section, I give an overview of the background and use of all Estonian pronouns as well as more specifically the three pronouns that are the main object of this research. Section 2.1 describes the Estonian pronoun system and lists studies that have been done on these pronouns. Sections 2.2, 2.3, 2.4 and 2.5 address the historical development, the possible dialect forms, the animacy reference and the grammaticalisation of *kes*, *mis* and *keegi*, respectively.

Pronouns are a closed set of words that can be used to substitute for a noun phrase. The exact reference or semantic interpretation of pronouns becomes evident only in context, as pronouns can be interpreted as picking out the same kinds of objects as full lexical nominals, but on their own, they lack descriptive content. When identifying the referent to which the pronoun is referring, one can rely on both discourse-pragmatic strategies and morphosemantic features. These features can tell us whether the referent is one vs. many, male vs. female, human vs. non-human, animate vs. inanimate, known vs. unknown, etc., and these can be linked to the referent by agreement in person, number, gender, etc. (Wiese & Simon 2002: 2–4, 9; Kroeger 2005: 136; Crystal 2008: 391).

There are many types of pronouns in the world's languages, with categories and terminology varying across grammars and languages. The most common ones, however, are personal pronouns, possessive pronouns, demonstrative pronouns, interrogative pronouns, reflexive pronouns, indefinite pronouns and relative pronouns (Crystal 2008: 391–392).

It is argued that the traditional definition of pronouns as words that stand for nouns is problematic, as many words that are typically categorised as pronouns can also substitute for adjective phrases, verb phrases and adverbs, sometimes even whole clauses or sentences. Instead, it has been suggested to divide the pronouns into personal pronouns and proforms, which encompass all words that substitute for other words or constructions (Bhat 2004: 272–273).

2.1. Estonian pronouns

In Estonian, pronouns morphologically resemble nouns, adjectives and numerals, using the same cases and numbers. Syntactically, they have the same grammatical relations as nouns (Alvre 1977: 18; Erelt et al. 1993a: 26). There are a few different ways to categorise Estonian pronouns, but most commonly they are separated into six types (Erelt 2000: 7):

1. personal pronouns,
2. demonstrative pronouns,
3. reflexive pronouns,
4. reciprocal pronouns,
5. interrogative-relative pronouns,
6. indefinite pronouns.

Personal pronouns represent the speaker, the listener or a third party. They are *mina ~ ma* ‘I’, *sina ~ sa* ‘you’, *tema ~ ta* ‘s/he’, *meie ~ me* ‘we’, *teie ~ te* ‘you:PL’, *nemad ~ nad* ‘they’ (Erelt et al. 1993a: 27). Estonian personal pronouns are notable for having both long and short forms, and many researchers have described and compared those forms or analysed their variation in both spoken and written language (see, e.g., Ariste 1956; Pajusalu 1997a, 1997b, 2005; Pool 1999; Kaiser 2003, 2010; Kaiser & Hiietam 2003). According to Standard Estonian, *tema ~ ta* is used to refer to animate entities, but in spoken Estonian, in the case of a very familiar entity or in some idiomatic expressions, the short form *ta* is often used to refer to inanimates (R. Pajusalu 2005: 132–133). Many studies have examined the way speakers alternate between informal *sina* and formal, polite *teie*, both in Estonian and in comparison with other languages (Keevallik 1999; Pajusalu 2010, 2014, 2019; Pajusalu et al. 2010a, 2010b, 2017). Vija (2007) has focused on the acquisition of the 1st and 2nd person pronouns by children and Kivik (2010) has examined the variation of personal pronouns in the speech of Estonians living in the United States. Several studies have been done on the ellipsis of the personal pronouns, showing that the ellipsis of different pronouns occurs on different terms (Keevallik 2003; Lindström, Kalmus, et al. 2009; Hint 2015; see also Lindström 2000, 2001a).

Demonstrative pronouns refer to a context-derived thing, characteristic or activity. Many demonstrative pronouns can be used both as prosubstantives and proadjectives. Demonstrative pronouns include *see* ‘it, this’, *too* ‘it, that’, *sama* ‘same’, *seesama* ‘exact same’, *teine* ‘another’, *muu* ‘other’, *niisugune* ‘this kind’, *samasugune* ‘same kind’, *selline* ‘this kind’, *taoline* ‘that kind’, etc. (Erelt et al. 1993a: 29). The demonstratives *see*, *too*, the dialect equivalent *taa* ‘it, that’ and their complex use has been studied extensively both in contemporary Estonian and in the southern Estonian dialects (Pajusalu 1996, 1997a, 1997b, 1998, 2005, 2006a, 2009, 2015; Reile 2016; Plado & Reile 2020; Reile et al. 2020, 2022). Although prototypically *see* refers to inanimate entities, Pajusalu (2005: 132–133) has shown that *see* can be used to refer to animate entities in a presentational clause, when there are two different animate referents involved or for emphasis in the case of a relatively new referent. Keevallik (2010) has described the functions of the pronoun *see* as a placeholder and Sahkai (2003) analyses the demonstrative-lexical phrases containing *see* as doubling constructions, which function as an important highlighting device in spoken Estonian. The use of demonstrative pronouns as (article-like) determiners is described and analysed in Pajusalu (1997c, 2009), Hint, Nahkola and Pajusalu (2017) and Hint et al. (2021). Many studies have addressed the referential properties of both personal and demonstrative pronouns (Tirkkonen 2007; Pajusalu 2009; Kaiser & Vihman 2010; Hint, Reile & Pajusalu 2013; Pajusalu et al. 2018; Reile et al. 2019; Hint, Nahkola & Pajusalu 2020; Hint 2021).

The aim of reflexive pronouns is to either emphasise that the activity is directed towards the actor themselves or that the possessor is co-referential with the subject. Reflexive pronouns are *ise* ‘oneself’, *oma* ‘one’s own’, *(ise)enese ~ (ise)enda* ‘one’s own’, *omaenese ~ omaenda* ‘one’s own’. Reciprocal pronouns

express a mutual relationship between two or more subjects, showing that they are acting in a similar way towards each other. There are two reciprocal pronouns: *üksteise* ‘one another’ and *teineteise* ‘each other’ (Erelt et al. 1993a: 28). Reflexive and reciprocal pronouns have been little researched in Estonian. Viks (1973), Erelt (1997), Õim (2001), and Lindström and Vihman (2010) have analysed the use of the pronouns *oma*, *ise* and *enese* ~ *enda*. In addition, Hint, Nahkola and Pajusalu (2017: 94–95) briefly describe the use of *oma* in possessive NPs.

Interrogative-relative pronouns have two main functions in discourse. One is to act as question words in (indirect) questions; the other is to function as a connecting word in relative clauses, where they substitute a NP. Interrogative-relative words include *kes* ‘who’, *mis* ‘what’, *kumb* ‘which’, *missugune* ‘what kind’, *milline* ‘what kind’, *mitu* ‘how many’, etc. (Erelt et al. 1993a: 29). In his articles, Alvre (1977, 1987a) has thoroughly described the various forms of *kes* and *mis*, both in contemporary Estonian and in Estonian dialects. Erelt (1996, 2004) has given an overview of relative clauses and the use of relative pronouns in those clauses in terms of animacy and agreement in number and person. Laanesoo (2013) has studied *mis*- and *mida*-initial questions in spoken language, while Pajusalu (2006b) has discussed the relative *wh*-constructions in written Estonian. The current dissertation is also an important contribution to the study of interrogative-relative pronouns.

Indefinite pronouns are a very diverse and large group of pronouns, sharing the characteristic of referring to something or someone unknown (Erelt et al. 1993a: 31). One of the most productive affixes in deriving indefinite pronouns in Estonian is *-gi/ki*, which functions in a similar way to discourse particles and has various meanings related to information structuring, quantification, etc. (Metslang 2003). Pronouns derived with this affix are *keegi* ‘someone’, *miski* ‘something’, *mingi* ‘some kind’, *ükski* ‘anyone’, *kumbki* ‘(n)either’, etc. In contemporary language, the affix *-gi/ki* has both additive (‘also’) and scalar (‘even’) meanings, but the original meaning of it is unclear. In negative contexts it behaves as a negative polarity item, as many words with this affix are used only with negative polarity (Sang 1983: 121–122; Paldre 1998: 49–51). Other indefinite pronouns include, but are not limited to, *iga* ‘every’, *igaiüks* ‘everyone’, *kõik* ‘all’, *mõlemad* ‘both’, *kogu* ‘whole’, *terve* ‘whole’, *mitu* ‘several’, *mõni* ‘some’, *paljud* ‘many’, *üks* ‘one’, *teine* ‘another’, etc (Erelt et al. 1993a: 31). Alvre (1995) has described etymology and the use of the pronoun *mõni* in both contemporary Estonian and in dialects. The use of the pronouns *mingi*, *üks* and *kõik* has been more thoroughly examined by Pajusalu (2000, 2001, 2004, 2008). For example, she has shown that while both *mingi* and *üks* express vagueness in spoken language, *mingi* is more likely to be used when the referenced entity is unfamiliar to both the speaker and the listener, and sometimes has a negative or an evaluative connotation, while *üks* conveys the meaning that in a given context the referent is unknown only to the listener (Pajusalu 2000). *üks* can also function as an article-like determiner (Hint, Nahkola & Pajusalu 2017). The variation of case forms in the pronouns *keegi*, *miski* and *kumbki* has been studied by Pant (2018, 2020).

2.2. The pronouns *kes*, *mis* and *keegi* from a historical perspective

In studying the history of the Finno-Ugric languages, it has been found that personal, interrogative and demonstrative pronouns are the oldest pronoun types in the Finno-Ugric languages, with many stems of these pronouns being shared even with the Samoyed languages. The relative and indefinite pronouns developed later, although in many Finno-Ugric languages the words used for interrogative and relative pronouns overlap (Alvre 1980: 539, 1986: 5, 1987b: 23).

The old pronoun stems consisted of one syllable: the interrogative *kes* developed from **ke*, the interrogative *mis* from **mi*. These pronouns have remained monosyllabic in all Finnic languages, but not always throughout of the entire paradigm. Both *mis* and *kes* most likely emerged from the Finnic proto-language, as Estonian shares these forms with Livonian. The source of the *s*-element in these pronouns has been theorised to be *se(e)* ‘it’, *sa* ‘you’ or *es* ‘not’, of which the first seems to be the most likely candidate (*mi-see* > *mi-s*) (Alvre 1977: 19, 1986: 6, 1987a: 28; Metsmägi, Sedrik & Soosaar 2012).

Initially, the basis for the oblique case forms of *mis* and *kes* were the unaffixed nominative stems, i.e., *ke-* and *mi-*. This means that the paradigms for the other cases used to be *ke-da*, *ke-l*, *ke-lle*, *ke-lt*, etc. and *mi-da*, *mi-l*, *mi-lle*, *mi-lt*, etc. The newer, reanalysed stem originated first from the allative case, to which another case ending was added (*ke-lle* > *ke-lle-le*, *mi-lle* > *mi-lle-le*). Later, the *-lle-* syllable began to be perceived as a part of the new stem and it spread to the entire paradigm (with the exception of the nominative and partitive). Old literary Estonian shows that this new stem evolved during the 17th–18th century (Ariste 1958: 687; Alvre 1977: 25, 1987a: 30–34; Prillop & Saar 2020: 303).

The late development of the indefinite pronouns is evidenced by their varied backgrounds and the existence of compound forms. The indefinite *keegi* derived from adding the **kin* ~ **kik* suffix to the interrogative stem. Deriving indefinites from interrogatives is typologically very common and it is characteristic of all the Uralic languages. In addition, the *gi*-suffix is one of the most productive affixes for deriving indefinite pronouns in the Finnic languages (Alvre 1980: 539, 1986: 10; Haspelmath 2013; Van Alsenoy & van der Auwera 2015).

The late development of the interrogative *keegi* can also be seen from the position of *-gi/ki*. As an enclitic particle, it is attached to the very end of the word after the number and case markers (*ilusa-te-le-gi* ‘beautiful-PL-ALL-CLI’), but as an affix on indefinites its position can vary, e.g., *kelle-le-gi* ~ *kelle-gi-le*. This positional variation is an indicator of the lexicalisation process, whereby the *-gi/ki* clitic has been recategorised as belonging to the stem and, therefore, its natural position is before the case and number suffixes (*kellegi-le*) (Alvre 1980: 540; Nevis 1984: 142–143).

Language planning has always suggested the placement of *-gi/ki* after other suffixes for *gi*-final pronouns, just as with other substantives. However, in dialects, spoken language and sometimes even written language, it is not uncommon to find *-gi/ki* before the case ending (Saareste 1933; see also Pant 2018, 2020).

For many dialects, the typical position of *-gi/ki* is before the case marker, at least in the allative form (Saareste 1955: 16). Moreover, *gi*-suffixed pronouns are not the only part of speech where *-gi/-ki* has ended up in the middle of the word: adverbs like *kusagil ~ kuskil* ‘somewhere’, *kuskile* ‘to somewhere’, *kusagilt ~ kuskilt* ‘from somewhere’ all exhibit the same phenomenon (Alvre 1980: 539–540).

2.3. The dialect forms of *kes*, *mis* and *keegi*

The articles in this dissertation have focused on the (morpho)syntactic variation of *kes*, *mis* and *keegi*, but attention was not given to the diverse variation of their forms in the dialects. Therefore, here I provide a brief summary of the various ways *kes*, *mis* and *keegi* are used in the dialects, based on different dialect overviews and dictionaries.

Tables 1, 2 and 3 illustrate the different nominative forms of *kes*, *mis* and *keegi* in all the Estonian dialects. These tables have been compiled based on many overviews of the pronouns and the dialects: Tanning (1961, 2004), Keem (1970), Alvre (1977, 1987a), Must (1987), Juhkam and Sepp (2000), Keem and Käsi (2002), Lonn and Niit (2002), Must and Univere (2002), Juhkam (2012) and Viikberg (2020). In addition, I have consulted the Estonian Dialect Dictionary (EMS II; EMS IIIa; EMS IIIb; EMS VI).

As can be seen from these tables, the possible forms for these pronouns are varied and diverse. Oftentimes, there are several different forms used in parallel in the same (sub)dialect. Some of these are used with similar frequency, while others represent older and newer forms of the same pronoun.

Table 1. Nominative forms of the pronoun *kes* in Estonian dialects

| | <i>kes</i> | <i>kis</i> | <i>ken</i> | <i>kie(s)</i> | <i>ke(e)</i> | <i>ket</i> | <i>kiä(s)/ keä</i> | <i>kens(e)</i> | <i>kiese</i> | <i>keese</i> | <i>kesse</i> | <i>kessi</i> | <i>kisse</i> | <i>kissi</i> | <i>ketse/ ketsi</i> | <i>kitse/ kitsi</i> |
|--------------|------------|------------|------------|---------------|--------------|------------|------------------------|----------------|--------------|--------------|--------------|--------------|--------------|--------------|-------------------------|-------------------------|
| Northeastern | new | | + | + | + | | | + | + | + | | | | | | |
| Coastal | new | + | + | + | + | | | + | + | | | | | | | |
| Insular | + | + | + | | + | old | | | | | | rare | + | rare | | rare |
| Western | + | + | | | | | | | | | | + | + | + | | |
| Mid | + | + | | rare | rare | | | | | | | rare | + | + | | |
| Eastern | + | rare | | | | | | | | | rare | + | | rare | | |
| Mulgi | new | new | + | | | rare | | | | | new | new | | | new | |
| Tartu | new | new | | | | | + | | | | | new | | | | |
| Võru | new | new | | | | | + | | | | | | | | | |
| Seto | new | new | | | rare | | + | | | | | | | | | |

Table 2. Nominative forms of the pronoun *mis* in Estonian dialects

| | <i>mis</i> | <i>mes</i> | <i>mäs</i> | <i>mih</i> | <i>meh</i> | <i>mi(i)</i> | <i>mee</i> | <i>mia/ miä(s)</i> | <i>meä</i> | <i>mida/ miä</i> | <i>midäs</i> | <i>miga(s)/ migä(s)</i> | <i>mige</i> | <i>mihe</i> | <i>misse</i> | <i>missi</i> |
|--------------|------------|------------|------------|------------|------------|--------------|------------|------------------------|------------|----------------------|--------------|-----------------------------|-------------|-------------|--------------|--------------|
| Northeastern | + | + | | | | + | rare | rare | | old | + | old | | | rare | rare |
| Coastal | + | + | | | | | | | | + | + | old | old | rare | | |
| Insular | + | + | | rare | rare | rare | rare | | | | | | | | + | + |
| Western | + | + | | | | | | | | | | | | | + | + |
| Mid | + | rare | | | | | | | | | | | | | + | + |
| Eastern | + | + | | | + | rare | | | | | | | | | | |
| Mulgi | + | new | | | | + | | | | | + | | | | | |
| Tartu | new | new | | | + | | | + | | | + | rare | | | | |
| Võru | new | new | | | | | | + | + | | + | | | | | |
| Seto | new | | | | | | | + | | | + | | | | | |

Table 3. Nominative forms of the pronoun *keegi* in Estonian dialects. *In the Võru and Seto dialects, the indefinite pronoun may lack the affix *-gi* (see Alvre 1977: 21), so the forms for *kes* and *keegi* can overlap.

| | <i>keegi</i> | <i>kiegi</i> | <i>kiigi</i> | <i>keski</i> | <i>kiski</i> | <i>kengi</i> | <i>keagi/</i> <i>keägi</i> | <i>kiägi/</i> <i>kiäki</i> | <i>kiäke</i> | <i>kessegi</i> | <i>kedagi/</i> <i>kedägi</i> | <i>keskid/</i> <i>kiskid</i> | <i>kengi(d)</i> | <i>ken(n)egi</i> | <i>kemig(i)</i> |
|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-------------------------------|-------------------------------|--------------|----------------|---------------------------------|---------------------------------|-----------------|------------------|-----------------|
| Northeastern | | + | | + | | + | | | | | + | | old | rare | |
| Coastal | | + | | + | | + | | | | | + | | old | rare | |
| Insular | rare | | | | | | | | | | | + | + | | |
| Western | + | | | + | + | | rare | | | | | | | rare | |
| Mid | + | + | rare | + | + | | rare | | | | | rare | | rare | rare |
| Eastern | | + | + | + | | | | | | | | | | | |
| Mulgi | | | | new | | rare | | rare | rare | | | | | + | + |
| Tartu | | | | + | | | + | + | | rare | | | | | |
| Võru* | | | | rare | | | + | + | rare | rare | | | | | |
| Seto* | | | | | | | + | + | + | | | | | | |

Next, I will briefly describe the possible genitive and partitive forms of *kes*, *mis* and *keegi* in all the dialects. All the oblique cases are derived from the genitive stem, which is the reason I do not present the other case forms here. For easy comparison of the dialect forms to the contemporary language, the paradigms of these three pronouns in Standard Estonian are given in Table 4.

Table 4. The full paradigms of *kes*, *mis* and *keegi* in Standard Estonian

| | | | |
|-------------|-----------------|-----------------|------------------|
| nominative | <i>kes</i> | <i>mis</i> | <i>keegi</i> |
| genitive | <i>kelle</i> | <i>mille</i> | <i>kellegi</i> |
| partitive | <i>keda</i> | <i>mida</i> | <i>kedagi</i> |
| illative | <i>kellesse</i> | <i>millesse</i> | <i>kellesegi</i> |
| inessive | <i>kelles</i> | <i>milles</i> | <i>kelleski</i> |
| elative | <i>kellest</i> | <i>millest</i> | <i>kellestki</i> |
| allative | <i>kellele</i> | <i>millele</i> | <i>kellelegi</i> |
| adessive | <i>kellel</i> | <i>millel</i> | <i>kellelgi</i> |
| ablative | <i>kellelt</i> | <i>millelt</i> | <i>kelleltki</i> |
| translative | <i>kelleks</i> | <i>milleks</i> | <i>kellekski</i> |
| terminative | <i>kelleni</i> | <i>milleni</i> | <i>kellenigi</i> |
| essive | <i>kellena</i> | <i>millena</i> | <i>kellenagi</i> |
| abessive | <i>kelleta</i> | <i>milleta</i> | <i>ketletagi</i> |
| comitative | <i>kellega</i> | <i>millega</i> | <i>kellegagi</i> |

In both the Northeastern and Coastal dialects, the typical genitive stem for KES is *kene*, with *kelle* being a newer addition to the dialects. The partitive form is *keda* ~ *kedä* or *kenne* (Alvre 1977; Must 1987: 229; EMS II: 893, 989, 1036; Viikberg 2020: 171–172). For MIS, the genitive stem is either *min(n)e* ~ *mene* or *mi(i)*, but *mille* and *mise* are also sporadically used. The partitive is *mida* ~ *midä* (Alvre 1987a; Must 1987: 229–230; EMS VI: 50, 86). Genitive stems of KEEGI include *kellegi* and *ken(n)egi* (and *kengi* for the Kuusalu subdialect), the partitive is *kedagi* ~ *kedägi* (Must 1987: 230; EMS II: 897, 994–995; Viikberg 2020: 173).

In the Insular dialect, the genitive stem of KES is generally *kelle* ~ *kille*, but older forms like *ken(n)e* ~ *kin(n)e* and *kesse* are also used. The partitive is typically *keda* or *kida*, while *s(t)*-final forms like *kedas(t)* and *kidast* are not unusual either (Alvre 1977; EMS II: 989, 1036, 1060; Lonn & Niit 2002: 51; Viikberg 2020: 171–173). For MIS, the genitive stem is typically either *mille* or *mis(s)e* ~ *mes(s)e*, more seldom *minne* or *mee*. The subdialects on the island of Hiiumaa also use *mike* ~ *meke*. The partitive is *mida* ~ *meda* (Alvre 1987a; Lonn & Niit 2002: 51; EMS VI: 50, 86; Viikberg 2020: 171–173). The genitive of KEEGI is either *kellegi(d)* or *ken(n)egi(d)*, the partitive is *kedagi(d)* ~ *kedägi*, *kedaged* or *kedad*, less often *kidagid* (EMS II: 897, 994; Lonn & Niit 2002: 51; Viikberg 2020: 173).

The genitive of KES in the Western dialect is mainly *kelle*, but sporadically the *kene* stem is used for declination. The partitive form is generally *keda*, more seldom *kedäst* (Alvre 1977; EMS II: 1036; Viikberg 2020: 171–173). The genitive stem of MIS is mainly *mille* (and *melle* in the Hanila subdialect), although *misse* stem is not unusual either. The partitive is *mida* (Alvre 1987a; EMS VI: 86; Viikberg 2020: 171–173). The genitive of KEEGI is *kellegi*, the partitive form is *kedagi* ~ *kedäge*, sporadically *kedagist* or *kedagid* (Juhkam & Sepp 2000: 23; EMS II: 897; Viikberg 2020: 174).

The Mid dialect mainly uses *kelle* ~ *kille* for the genitive stem of KES, while sometimes *ken(n)e* occurs as well. The partitive form is *keda*, rarely *kedä* (EMS II: 1036; Must & Univere 2002: 271; Juhkam 2012: 468–469; Viikberg 2020: 171–173). For MIS, the genitive is *mille* and the partitive is *mida*, rarely *midä* (Must & Univere 2002: 271; Juhkam 2012: 468–469; EMS VI: 86; Viikberg 2020: 171–173). The genitive of KEEGI is mostly *kellegi*, although *kenegi* is also used in certain areas. The partitive is *kedagi* ~ *kedägi*, sometimes also *kedagid* and *kedagist*. Rarer partitive forms include *kedagitki* and *kenegit* (EMS II: 897, 994; Must & Univere 2002: 274; Juhkam 2012: 469; Viikberg 2020: 174).

In the Eastern dialect, the genitive stem of KES is *kelle* and the partitive is *keda* ~ *kedä* (EMS II: 1036). The genitive stem of MIS is *mille*, while the partitive is *mida*, rarely *medä* (EMS VI: 86). The genitive of KEEGI is *kellegi*, the partitive *kedagi* ~ *kedägi* (EMS II: 897).

In the Mulgi dialect, the typical genitive stem of KES is *kelle*. The partitive form is *kedä*, sporadically *kedäs* can also be used (Tanning 1961: 45, 2004: 110; Alvre 1977; EMS II: 989, 1036). For MIS, the genitive is *mille*, *minne* or *mike*. In some places, the *mi*-stem is also used in oblique case forms. The partitive is *midä* or *miket* (Tanning 1961: 45, 2004: 110; Alvre 1987a; EMS VI: 50, 86). The genitive stem of KEEGI is *kellegi* ~ *kelleki*, sometimes *kennegi/kennigi*, the partitive is *kedägi(t)* (Tanning 1961: 45, 2004: 111; EMS II: 897, 995; EMS IIIb: 302).

The typical genitive stem of KES in the Tartu dialect is *kelle*, but *kenka* is used as well. The partitive is *keda*, rarely also *kidä(d)* (Keem 1970: 40; Alvre 1977; EMS II: 1036; EMS IIIb 301). For MIS, the genitive stem is mainly *mille*, with some subdialects using *melle*, *mika* or *me* stems in oblique case forms. The partitive is *midä* or *medä*, rarely *midäd* (Keem 1970: 40; Alvre 1987a; Keem & Käsi 2002; EMS VI: 45, 86). The genitive of KEEGI is *kellegi* ~ *kelleki*, the partitive is *kedägi* or *kidägi* (EMS II: 897; EMS IIIb: 302).

The Võru dialect uses mainly *kink(a)* stem for the genitive of KES, more seldomly *kinga*, *kenne*, *kenka* or *kii*. Genitive *kelle* is a newer addition to the dialect. The partitive is *kedä* or *kidä* (Alvre 1977; EMS II: 1036; EMS IIIb: 301; Keem & Käsi 2002: 44). For MIS, the genitive stem is *mink(a)* or *minga*, sometimes *mii* ~ *mee*, with *mille* being a new form. The partitive is *midä*, sometimes *midäs*, rarely *medä* or *mitä* (Alvre 1987a; Keem & Käsi 2002: 44; EMS VI: 45, 86). The forms of KEEGI can overlap with those of KES, however, genitive *kellegi* ~ *kelleki* and partitive *kedägi* are also used (EMS IIIb: 302; Keem & Käsi 2002: 44).

In the Seto dialect, the genitive stem for KES is *kink(a)* or *kinkõ*, recently also *kelle*. The partitive forms are *kedä*, *kettä* or *kinkõt*, sporadically also *kedäs* (Alvre

1977; EMS II: 1036; EMS IIIb: 301). The genitive stem for MIS is *mink(õ)*, the partitive is *midä* or *midäs* (EMS VI: 45, 86). Similarly to the Võru dialect, the forms for KES and KEEGI sometimes overlap, but genitive *kinkõgi* and partitive *kedägi* are also used (EMS IIIb: 302).

In the Tartu, Võru and Seto dialects, it is not uncommon that while speakers use the nominative forms *kiä* and *kiägi*, for the rest of the cases the Standard Estonian declination of *kes* and *keegi* is used instead (EMS IIIb: 301).

As a comparison, I have compiled similar pronoun tables for the data from the Corpus of Estonian Dialects, i.e., the datasets that have been used in this dissertation. Corpus data do not, however, represent very archaic dialect speech, but rather a more levelled version of the dialects, since the data have been collected when the dialects were already under heavy influence from the standard language. This can immediately be seen in Tables 5, 6 and 7. Many forms that are considered to be new in certain dialects by the dialect overviews actually appear regularly in those dialects in the corpus dataset.

By the time of the data collection, all dialects frequently used the Standard Estonian *kes* and *mis*. *kis*, *mes*, and for the southern dialects *kiä*, *miä*, are also common, while other forms tend to be rarely or never used. There is also significantly less variation in the possible forms in the corpus data for these two pronouns. *keegi* has more variety and many of the forms described in the dialect overviews also appear in the corpus datasets. However, the same tendencies as with *kes* and *mis* can be seen: the use of the Standard Estonian *keegi* has spread into all the northern dialects, but is still rare or non-existent in the southern dialects. *kiegi*, *kiigi*, *keski* and *kiä(gi)* are frequently used as well, most other forms are rare.

Table 5. Nominative forms of the pronoun *kes* in the Corpus of Estonian Dialects

| | <i>kes</i> | <i>kis</i> | <i>ken</i> | <i>kie(s)</i> | <i>ke(e)</i> | <i>kiä(s)</i> | <i>kea/keä</i> | <i>kesse</i> | <i>kessi</i> | <i>kisse</i> | <i>kissi</i> |
|--------------|------------|------------|------------|---------------|--------------|---------------|----------------|--------------|--------------|--------------|--------------|
| Northeastern | + | rare | + | + | | | | rare | | | |
| Coastal | + | rare | rare | | | | | rare | | | |
| Insular | + | + | rare | | rare | | rare | | rare | rare | |
| Western | + | + | | rare | | | | | | rare | rare |
| Mid | + | + | | | | | | | | rare | rare |
| Eastern | + | + | | | | | | rare | | | |
| Mulgi | + | rare | | | | | | rare | | | |
| Tartu | + | rare | | | | rare | | | | | |
| Võru | + | + | | | | + | + | | | | |
| Seto | + | rare | | | | + | | | | | |

Table 6. Nominative forms of the pronoun *mis* in the Corpus of Estonian Dialects

| | <i>mis</i> | <i>mes</i> | <i>mäs</i> | <i>mih</i> | <i>meh</i> | <i>mi(i)</i> | <i>mia/miä(s)</i> | <i>meä</i> | <i>migä(s)</i> | <i>misse</i> |
|--------------|------------|------------|------------|------------|------------|--------------|-------------------|------------|----------------|--------------|
| Northeastern | + | rare | | | | | | | | |
| Coastal | + | + | | | | | | | + | rare |
| Insular | + | + | rare | | rare | | | | | |
| Western | + | + | | | | rare | | | | |
| Mid | + | + | | | | | | | | + |
| Eastern | + | + | | | | | | | | |
| Mulgi | + | + | | rare | | rare | rare | | | |
| Tartu | + | + | | | + | rare | | | | |
| Võru | + | rare | | | rare | rare | + | rare | | |
| Seto | + | rare | | | | | + | | | |

Table 7. Nominative forms of the pronoun *keegi* in the Corpus of Estonian Dialects.

| | <i>keegi(d)</i> | <i>kiegi(d)</i> | <i>kiigi</i> | <i>keski/ kiski(d)</i> | <i>kengi</i> | <i>keagi/ keägi</i> | <i>kiägi/ kiäki/ kiagi</i> | <i>kiäke/ keäke</i> | <i>kedagi</i> | <i>ken(n)egi</i> | <i>kennig(i)</i> | <i>kuagi</i> | <i>kes</i> | <i>kiä</i> |
|--------------|-----------------|-----------------|--------------|----------------------------|--------------|-------------------------|------------------------------------|-------------------------|---------------|------------------|------------------|--------------|------------|------------|
| Northeastern | rare | + | | + | rare | | | | | rare | | | | |
| Coastal | rare | + | | + | | | | | | rare | | | | |
| Insular | + | + | | rare | | | | | rare | | | | | |
| Western | + | | rare | | | rare | | | | | | | | |
| Mid | + | + | + | rare | | | + | | | | | | | |
| Eastern | + | rare | + | rare | | | | | | | | | | |
| Mulgi | | | | rare | | | + | | | rare | + | | rare | |
| Tartu | | | | rare | | | + | | | | | | rare | rare |
| Võru | | | rare | rare | | | + | rare | | | | rare | rare | + |
| Seto | rare | | | rare | | | + | | | | | | rare | + |

2.4. Animacy in language and the animacy reference of *kes*, *mis* and *keegi*

Animacy often plays an important role in syntactic variation. Indeed, it is a central concept also in this dissertation, as one of the main goals of the dissertation has been to analyse the animacy reference of the pronouns *kes*, *mis* and *keegi*. Therefore, in this section, I give an overview of the complexity of animacy in language and describe the ways that animacy determines the use of the studied pronouns *kes*, *mis* and *keegi*.

In language, animacy is rarely regarded as a binary variable, but rather as a continuum that extends from human through animal to inanimate and that is not necessarily always aligned with biological animacy (Yamamoto 1999: 1; Bayanati & Toivonen 2019: 158–159). The first modern description of this continuum was presented by Silverstein with regard to ergative languages (Silverstein 1976), but the most common representation of it, called the animacy hierarchy, is found in Dixon (1979: 85):

1st, 2nd person pronoun > 3rd person pronoun > proper names > human common noun > non-human animate common noun > inanimate common noun

The existence of some kind of animacy hierarchy has been claimed to be linguistically universal, however, not all languages or grammatical structures associated with animacy distinguish between all the levels of the hierarchy. Some languages and/or phenomena use less fine distinctions, e.g., just distinguishing humans from non-humans or animate beings from inanimate beings. Others can have intermediate categories between the levels of the base hierarchy, e.g., separating higher animals from lower animals or having 1st and 2nd person as separate categories (Dixon 1979: 85; Comrie 1989: 185; Croft 1990: 113; Bayanati & Toivonen 2019: 157).

It has been shown cross-linguistically that the categories higher in the hierarchy are often grammatically distinguished from those that are lower, are treated as more central to clause structure and are more likely to act as an agent in events. Linguistic animacy can condition a wide range of phenomena, e.g., case, agreement, number, word order, voice selection, form of reference, argument structure patterns, classifiers, etc. (Whaley 1996: 172; Kittilä, Västi & Ylikoski 2011: 6; Vihman & Nelson 2019: 261). Nevertheless, this hierarchy is not really an ordering of discrete categories, but rather a continuum ranging from most to least animate (Croft 1990: 113).

In addition to animacy, there are other important parameters reflected in this hierarchy like empathy, individuation, definiteness, focus, grammatical relations, semantic roles, etc. For example, the hierarchy considers the NPs with which the speaker is familiar or toward which the speaker has empathy to be more dominant. Likewise, a stronger sense of animacy is attributed to the entities that are seen as individuals rather than a part of a group or an indeterminate mass. Therefore, singular NPs can be considered higher on the animacy hierarchy than plural NPs,

but there are also several examples of groups that lie on the boundary between animacy and inanimacy, e.g., human organisations, geographic entities or communities. Most of these are usually treated as inanimate, but they still show certain animate-like features (Comrie 1989: 189; Whaley 1996: 172–173; Yamamoto 1999: 131, 138–139; Dahl 2008: 142).

Some authors prefer to avoid the term *animacy* altogether and speak rather of empathy hierarchies, since there is technically no difference in literal animacy between the 1st person and a proper noun or a human common noun (Kittilä, Västi & Ylikoski 2011: 6; see also Langacker 1991: 306–307). All in all, determining whether something in language is animate or inanimate is not always clear. Not even all animate beings are equally animate to human cognition, since animacy is on a gradient from centrally animate beings to peripherally animate beings (Yamamoto 1999: 14).

Recent decades have seen a surge of interest in animacy and its role in the world's languages, and consequently several volumes and special issues on animacy in language have been published (see, e.g., de Swart, Lamers & Lestrade 2008; Lamers & de Swart 2012; Gardelle & Sorlin 2018; Vihman & Nelson 2019). All these and other studies have shown that the role of animacy in the grammars of the world's languages is extremely diverse and there exist a wide variety of ways animacy can affect the syntax and semantics of a language.

The research objects of this dissertation – the pronouns *kes*, *mis* and *keegi* – are an interesting example of how animacy can determine the way that speakers use certain syntactic constructions. In most Finno-Ugric languages, the distinction between animate and inanimate is made with contrasting pronoun stems, as opposed to many Indo-European languages, which typically use suffixes to distinguish between singular and plural or animate and inanimate pronouns (Alvre 1977: 18). The choice between the Estonian interrogative-relative pronouns *kes* 'who' and *mis* 'what', and the indefinite pronouns *keegi* 'someone, no one' and *miski* 'something, nothing' in Standard Estonian is determined by the animacy hierarchy and the individuation scale (Erelt 2017: 743):

humans > higher animals > lower animals > inanimates
an individual > a distributive group > a collective

With these hierarchies, the probability of using *kes* and *keegi* decreases and the probability of using *mis* and *miski* increases from left to right. In Standard Estonian, the relative pronoun *kes* is always used for a single human referent and the relative pronoun *mis* is always used to refer to an inanimate entity. In the case of multiple referents, *kes* is preferred for a distributive group (wherein for a group of humans, only the use of *kes* is possible) and *mis* is preferred for a collective group. The choice of the pronoun also depends on whether the collective or institution is in the role of agent or patient – in the first case, *kes* is preferred, while in the second case, *mis* is preferred. In general, the more active the role of the referenced entity, the more likely that the pronoun *kes* will be used (Erelt 2014: 743; see also Veski 1958: 31).

Animate non-human entities can be referred to using both *kes* and *mis*. The pronoun *mis* can even be used to refer to a single animal: with higher animals, the use of *mis* is justified if the animal is not the agent in the sentence, for lower animals the choice of the pronoun is freer and it is always possible to use the pronoun *mis* (Erelt, Erelt & Ross 2007: 561). If either *kes* or *mis* can be used according to the hierarchies, then the choice of the pronoun is determined by other factors, mostly the meaning of the verb and the semantic function of the argument (Erelt 1996: 13).

However, in some Estonian dialects the pronouns *kes* and *keegi* can be used to refer to inanimate entities and the pronoun *mis* (although not *miski*) can refer to animate entities. Several linguists have brought up this possibility in their dialect overviews. Kask and Palmeos (1985: 59) mention that in some areas of Estonia the relative pronouns *kes* and *mis* can be used opposite to the way they are used in Standard Estonian. Alvre (1977: 19, 23, 1987a: 32) points out that using *kes* in the meaning of ‘what’ is quite common in the northern dialects, and that the meaning ‘who’ is sometimes attributed to both *mida* in the northeastern dialects and to *koa ~ kua* ‘which’ in the southern dialects. Tanning (2004: 112) and Viikberg (2020: 174) state that *midägi* – the partitive form of the inanimate *miski* – is extremely rare in the Mulgi dialect and that the animate *kedägi* is used instead.

Wiedemann (2011 [1875]: 470–471) notes in his Estonian grammar that in central Estonia only the interrogative pronoun *kes* (referring to animates) was carefully kept apart from the interrogative pronoun *mis* (referring to inanimates), but while using these pronouns in a relative clause, a similar distinction was often not made. However, in some dialects, this distinction was not made for interrogative pronouns either. In addition, he mentions that if the declension system was missing a certain case for the pronoun *mis*, it was replaced by the pronoun *kes* in the respective case. This phenomenon of using animate pronouns to refer to inanimate entities and vice versa is also noted in a few descriptions of certain Estonian (sub)dialects (see, e.g., Niklus 1957: 139; Tekko 1958: 110; Pung 1968: 134; Lonn & Niit 2002: 55; Must & Univere 2002: 271; Juhkam 2012: 468). Nevertheless, all of these previous works only mention the existence of this possibility, but do not elaborate on how or why *kes*, *mis* and *keegi* are used in this way.

Examples that run contrary to the principles of Standard Estonian have not only been found in Estonian dialects, but are also apparent in old literary Estonian. While in those texts *kes* usually refers to animate entities, there are certain relative clauses where the relative pronoun *kes* refers to inanimate referents, such as *usk* ‘belief’, *puu* ‘tree’, *kiri* ‘letter’, etc. (Kõpp 2001: 41–42). There are also cases where the relative pronoun *mis* is used to refer to animate referents, e.g., *laps* ‘child’, *mina* ‘I’, *loomad* ‘animals’, but those are less frequent (Kõpp 2001: 67–68). A similar phenomenon can be seen in many 17th and 18th century texts, and this kind of opposite use of the pronouns has extended to even some written texts from the 19th and early 20th century (Alvre 1977: 19; Erelt 1996: 11–12).

All this shows that the choice between the pronouns *kes* and *mis* and *keegi* and *miski* has not always functioned according to contemporary norms of Standard Estonian. Interestingly enough, the clear distinction of animacy between *kes* and *mis* was designated in a language planning meeting in 1911, making it one of the first language planning decisions made for the Estonian language (Veski 1958: 31).

kes, *mis* and *keegi* are not the only Estonian pronouns that have lost some of their distinction in animacy in spoken language; a similar phenomenon can be observed for the pronouns *tema* ‘s/he’ and *see* ‘it’. Typically the demonstrative pronoun *see* is used to refer to inanimate entities and the personal pronoun *tema* is used to refer to animate entities. However, in spoken Estonian, in the case of a very familiar entity or in some idiomatic expressions, *ta* (the short form of the pronoun *tema*) is often used to refer to inanimates. Similarly, the demonstrative *see* is used to refer to animate entities either in a presentational clause, when there are two different animate referents involved or for emphasis in the case of a relatively new referent. This kind of use can also indicate a negative attitude towards the referent. Nevertheless, if there are two referents, one of which is animate and the other inanimate, the two pronouns are used in their prototypical way to distinguish between the referents (Pajusalu 2005: 110, 132–133, 2017: 576).

In some dialects, the choice of some demonstrative pronouns is also based on the animacy of their referent. For example, in the Võru dialect the pronouns *timä* or *ta* ‘s/he’ are used to refer to animate referents, but the pronoun *tuu* ‘that’ typically refers to an inanimate or even abstract entity. However, distinguishing between these two pronouns is not always clear-cut (Pajusalu 2015: 173).

In addition to affecting pronouns, animacy can determine the use of many other morphosyntactic constructions in Estonian (a few examples are given in Pook (2018: 12–13)), all of which show that animacy in language is complex and often scalar in character. Nevertheless, in this dissertation I have chosen to follow Fowler’s (1977: 16–17) distinction for animacy, which divides all entities into two groups: animate and inanimate. Animate beings are those that are capable of movement (which is why plants are typically perceived as inanimate) and of initiating actions and change, either through conscious or unconscious action (whereas inanimate beings lack the ability to cause the world to change). Therefore, all humans (and human collectives) as well as animals are marked as animate and everything else as inanimate. The choice to view animacy as binary in my dissertation stems from the data used: due to the everyday topics of the dialect corpus recordings, there are rarely any borderline cases for animacy and all NPs can easily be categorised according to their biological animacy. Moreover, since this dissertation studies the animacy of pronouns, many of the finer categories on the animacy hierarchy cannot be applied to them at all.

2.5. The grammaticalisation of *kes*, *mis* and *keegi*

Grammaticalisation can be an important factor behind language change. It is proposed in this dissertation that grammaticalisation might also be one of the driving forces in the development of the variation in the animacy reference of the pronouns *kes*, *mis* and *keegi*. Therefore, in this section, I give an overview of grammaticalisation as a theory and explain how grammaticalisation is connected to the use of the three studied pronouns.

Grammaticalisation is a process where certain language forms develop from lexical to grammatical or from grammatical to more grammatical. This process can include phonemes and words, but also constructions and larger discourse units (Lehmann 1985: 303; Heine & Kuteva 2002: 2). The change does not take place abruptly, but instead the forms gradually go through a series of small transformations. This is called a cline of grammaticality and the following is a general representation of it (Hopper & Traugott 2003: 6–7; see also Lichtenberk 1991):

content item > grammatical word > clitic > inflectional affix.

Probably the most known example of grammaticalisation in Estonian (and also in several other Finnic languages) is the development of the comitative case. The function of co-existence and co-occurrence was originally filled with the noun *kansa* ‘with, together’. The lative form of this noun, **kansak*, began to be used as a postposition, and, in turn, went through a number of phonetic changes, resulting in the form *kaas*, which is already present in old literary Estonian. Going forward, after the loss of the genitive ending *-n*, it became easier to attach the postposition to head nouns. This unstressed position gave rise to multiple irregular changes: the postposition lost the word-final *s*, the intervocalic fortis stop *k* was replaced by the lenis stop *g* and thus the modern comitative ending was born. The entire cline of grammaticalisation for the comitative case is as follows (Tauli 1966: 112–113; Rätsep 1979: 77–78; see also Campbell 1991):

**kansak* > **kaasak* > **kaasa* > *kaas* > *ka(a)* > *-ga*

Grammaticalisation is typically viewed as a unidirectional process, with very few contradictory examples. However, there is no fixed end for grammaticalisation, as the process can stop at any point of development, and even in the final stage, a form might still be in variation and earlier forms may coexist with newer ones (Lichtenberk 1991: 37; Heine & Kuteva 2002: 4–5; Hopper & Traugott 2003: 16, 67, 127, 138).

Over time, linguists have proposed many principles and parameters for describing and identifying grammaticalisation (see, e.g., Hopper 1991). Heine and Kuteva (2002: 2) suggest the following four (interrelated) parameters of grammaticalisation:

- desemanticisation or semantic bleaching (i.e., the forms lose some of their meaning),

- extension or context generalisation (i.e., the forms are used in new contexts),
- decategorialisation (i.e., the forms lose morphosyntactic properties that are characteristic of either lexical or less grammaticalised forms),
- erosion or phonetic reduction (i.e., the forms lose phonetic substance).

Frequency also plays an important role in grammaticalisation. Some authors believe that a high frequency of occurrence is what makes a form eligible for grammaticalisation in the first place, while some consider a (sudden) increase in the (proportional) discourse frequency a sign that grammaticalisation has already occurred. Many changes characteristic of grammaticalisation (like semantic bleaching, erosion, positional fixing, erasure of word boundaries) are directly related to the absolute frequency of the forms; for example, as lexical forms acquire new grammatical meanings, they tend to be used more frequently, which directly contributes to their phonetic reduction (Heine, Claudi & Hünne-meyer 1991: 38–39; Heine & Kuteva 2002: 3; 2005: 45–46; Hopper & Traugott 2003: 127; Mair 2004: 125). Nevertheless, frequency alone cannot be considered to be the driving force behind the grammaticalised changes (Heine, Claudi & Hünne-meyer 1991: 39; Hoffmann 2004: 189).

It is proposed in this dissertation that grammaticalisation has, at least in part, influenced the use of the pronouns *kes*, *mis* and *keegi* studied in this dissertation. According to Heine and Kuteva (2006: 209), interrogative markers have four possible stages of grammaticalisation, from being just an interrogative marker to introducing headed relative clauses:

1. Only an interrogative marker (*Who came?*).
2. The interrogative marker is extended to introduce indefinite complements or adverbial clauses (*I don't know who came.*).
3. In addition to indefinite clauses, the interrogative marker can introduce a definite complement or adverbial clauses, which can also be interpreted as headless relative clauses (*You also know who came.*).
4. In addition to definite non-headed relative clauses, the interrogative marker can introduce headed relative clauses (*Do you know the woman who came?*).

It has been noted for many languages that going through these stages decreases distinctions in gender, animacy, number and case, and that the marker assumes a more general (grammatical) meaning (Heine & Kuteva 2006: 227). Lindström (2006) has proposed that this cline of grammaticalisation has also played a role in the change of word order in subordinate clauses, since grammaticalised relative pronouns are typically not focused anymore.

The Estonian interrogative markers *kes* and *mis* have gone through all of these grammaticalisation stages (Erelt 2017: 683), and indeed, it is shown also in this dissertation that *kes* is most likely to refer to an inanimate entity and *mis* is most likely to refer to an animate entity precisely in their most grammaticalised position

(the relative clause), while the least grammaticalised position (the interrogative clause) allows for considerably less variation in terms of the animacy reference of these two pronouns. This indicates that grammaticalisation has played an important role in the development of non-selectivity of *kes* and *mis*.

In addition, it is proposed that the pronoun *keegi* might also be in the process of grammaticalisation, as the loss of semantic distinctions and semantic bleaching in general can be an indication of an early stage of grammaticalisation. This is further supported by the results of P2, where it was found that *keegi* is most likely to refer to an inanimate entity in its most grammaticalised function: the negative polarity item. This grammaticalisation, however, seems to be occurring mostly in the scope of negation and the animacy distinction of *keegi* is not that much affected in those syntactic positions for which differentiating between animate and inanimate referents is still relevant to understanding the content of the clause (e.g., nominative subjects). Further discussion about the effect of grammaticalisation on the use of *kes*, *mis* and *keegi* can be found in P1 and P2.

3. DIALECTOLOGY AND SYNTACTIC VARIATION

This dissertation is a study of dialectology, the aim of which is to examine the syntactic variation of the pronouns *kes*, *mis* and *keegi* in Estonian dialects. In Section 3.1, I give an overview of dialectology and its development into an academic linguistic field, both in the world and in Estonia. In addition, I explain the concept and benefits of corpus-based dialectology, which is the approach used in this dissertation, and introduce the framework of dialect syntax. Section 3.2 describes the study of (syntactic) variation and its methods.

3.1. Dialectology

Dialectology, the study of dialects, started its modern tradition in Europe in the 19th century. The earliest dialect collectors were parish priests and schoolteachers, many of whom were motivated by their concern to record as much traditional rural speech as possible before it was altered or lost to industrialisation and urbanisation. Over time, however, the study of dialects started to develop a more academic side (Boberg, Nerbonne & Watt 2018: 5–6).

The first survey that could be interpreted as belonging to academic dialectology was conducted during 1876–1887 by Georg Wenker, who asked teachers all over Germany to translate a set of 40 sentences into their local dialect. The French linguist Jules Gilliéron soon followed with his own dialect survey. In order to collect his data, however, Gilliéron used trained fieldworkers, who were sent to different regions to conduct interviews and record data in a consistent phonetic notation (Chambers & Trudgill 1998: 15–16; Boberg, Nerbonne & Watt 2018: 6).

The goal of these two projects (and the many that followed them) was to produce dialect atlases: collections of maps showing the regional variation of different linguistic phenomena. Mostly these maps demonstrated the use of alternate words, pronunciations and grammatical forms. Based on their surveys, Wenker and Gilliéron published two of the first dialect atlases, *Deutscher Sprachatlas* and *Atlas linguistique de la France* (Chambers & Trudgill 1998: 15; Boberg, Nerbonne & Watt 2018: 6).

Out of the first dialect surveys grew the traditional method of collecting dialect data via questionnaire-based fieldwork with native-speaker informants. This fieldwork was typically distributed evenly across all dialect areas with a large number of measurement points. Up until the end of the last century, the questionnaires focused on lexical and phonetic variation, sometimes including a few morphological phenomena, but syntactic information was not yet gathered (Kortmann 2010: 843).

A newer method in dialectology for collection data is the typology-style analytic questionnaire, which is completed by specialist for individual dialects or, if possible, native speakers themselves. These questionnaires collect information on whether particular phenomena are prevalent or at least observed in the dialects,

or whether they are not part of certain dialects at all. They can also shed light on features that have otherwise been overlooked in studies on individual dialects, either because they are not used frequently enough, they are not distinctive (enough) or they have been considered a part of spontaneous spoken language in general. Such questionnaires help to specify the areal spread of the analysed features (Kortmann 2010: 843–844).

In recent decades dialectometry has become an increasingly popular approach in dialectology. Dialectometry is a method that assesses the linguistic distance between selected dialect areas or points. While traditional dialectology often focuses on one phenomenon, either in one specific dialect area or in all dialects of a particular language, dialectometry aims to identify more general structures from a large amount of features, in order not to characterise a single feature, but rather the dialects themselves and the relations between them (Nerbonne & Kleiweg 2007: 148; Szmrecsanyi 2013: 2–3). Dialectometry is considered to be established by Séguy (1971; 1973) and was later much developed by Goebel (see, e.g., 1982, 2006, 2011). The most common fields of research in dialectometry are phonetics and lexis, although a few (morpho)syntactic studies have been conducted as well (Szmrecsanyi 2013: 2).

3.1.1. Corpus-based dialectology

The introduction of different speech recorders and the subsequent emergence of many dialect corpora in the 20th century have paved way for variety of new directions in dialectology, one of them being corpus-based dialectology. Corpus linguistics is a methodology that uses systematic collections of texts (corpora) to analyse and make claims about linguistic phenomena and linguistic variation. Corpora can contain very different material: written and spoken language, modern and historical language, standard and non-standard language, adult or child language, native or learner language, etc. (Szmrecsanyi & Anderwald 2018: 300).

Large amounts of computerised material in corpora allow for conducting solid quantitative and qualitative analyses, especially for high-frequency phenomena. By analysing texts in a corpus, corpus linguists are first and foremost interested in what the speakers do with the language and not what they know (or think they know) about the language or their own language use (Kortmann 2010: 844–845; Szmrecsanyi & Anderwald 2018: 300).

Corpus linguistics can be divided into two approaches: corpus-based and corpus-driven. Corpus-based studies use the corpus data in order to explore an existing hypothesis or to analyse patterns of variations of those linguistic forms and structures that are already pre-defined. A corpus-driven approach, however, makes no prior assumptions about the data and claims that the corpus itself should be the only source of hypotheses about language (Biber 2009: 162; McEnery & Hardie 2011: 6). In this dissertation, I follow the corpus-based approach to conduct my research.

Dialect corpora contain natural (transcribed) discourse, which is based on interviews and/or free speech by dialect speakers. These transcriptions can be

either monologues or conversations between the informants or between the informant and the interviewer (Kortmann 2010: 844; Szmrecsanyi & Anderwald 2018: 301).

There are many advantages to using dialect corpora over linguistic atlases in order to study the dialects. The data in dialect atlases rely on elicitation and questionnaires, which makes it twice removed from the reality of language: firstly via fieldworkers, secondly via the atlas compilers. The answers by the speakers might also represent the conscious use of language rather than their real language use. Corpora allow us to access a more direct, spontaneous form of language, which, in turn, yields a more realistic and reliable result (Szmrecsanyi 2013: 3–4; Lindström & Pilvik 2018: 649).

As atlases include categorical data, they might omit borderline cases, while corpora provide graded frequency information that is more suitable for analysing continuous linguistic variation. Frequencies, in turn, represent the reality of language much better than the discrete data of the atlases (Szmrecsanyi 2013: 3–4). In addition, a corpus allows the researcher to analyse all structural levels of a language. Corpora not only include the use of a specific phenomenon, but also the immediate context, and this provides the option of researching the semantic, pragmatic, textual, etc. features as well. In practise, the actual research possibilities of course depend on how the particular corpus has been compiled and annotated (Lindström & Pilvik 2018: 649). Lastly, corpora typically include information about the informants (their age, birthplace, education, field of work, etc.), which allows researchers to include social parameters in their variation analysis (Lindström & Pilvik 2018: 649).

Certainly there are drawbacks to corpus-based research as well. A major shortcoming of corpora is the possible absence of rare phenomena. Some rare features might not be present in the corpus not because of their rarity in the dialect, but because of the rarity of the specific discourse contexts in the corpus. Similarly, even relatively frequent phenomena might not show up in the corpus in the particular contexts that the linguist would like to examine. This drawback especially affects syntax studies, as syntactic phenomena tend to be rarer than, for example, phonetic phenomena. Thus, in order to analyse dialect syntax, a linguist needs a lot more text than for other research fields, since only features that are frequent enough can be analysed. However, it has also been argued that if a phenomenon or a specific context is so rare it is not possible to describe it based on large text corpora, then it might not need an in-depth analysis at all (Anderwald & Szmrecsanyi 2009: 1136; Kortmann 2010: 845; Szmrecsanyi 2013: 4; Szmrecsanyi & Anderwald 2018: 308–309).

It is also clear that the material included in the corpus puts a range on the possible research questions that are answerable on the basis of it (Szmrecsanyi & Anderwald 2018: 300). Lastly, dialect corpora often have a smaller areal coverage than dialect atlases, as corpora tend to gather more material but from fewer locations (Szmrecsanyi 2013: 3).

All in all, questionnaire-based and corpus-based methods complement each other, allowing the linguist to collect data in different degrees of granularity.

Although the (analytic) questionnaires primarily give a crude picture of (syntactic) variation in particular regions, they can be very helpful in determining which phenomena are worth investigating further. Similarly, elicitation tasks and questionnaires are sometimes needed in addition to corpus data to determine the exact syntactic and pragmatic circumstances in which the rarer features in corpora are used in (Anderwald & Szmrecsanyi 2009: 1137; Kortmann 2010: 846; see also Rosenbach 2013). In this dissertation, however, additional data collection methods have not been used, as the dialects have severely levelled by the time this dissertation was written, compared to the 1960s and 1970s when the original material was collected. Fortunately, corpus data alone have been enough to conduct a comprehensive syntactic analysis.

3.1.2. Dialect syntax

Dialect variation in the field of syntax has until the recent decades been mostly neglected – the focus had either been on standard languages and their syntax or on phonological and morphological variation in dialects. The few early syntax studies focused mostly on single features in smaller dialect areas, while at the same time the research done in phonology and lexis often already included much larger areas and/or multiple features (Szmrecsanyi 2013: 1).

There are many reasons why dialect syntax was sidelined for a long time. For example, it was once thought in some dialectological communities that (morpho)-syntactic variation in dialects is not connected to geographic location at all (Szmrecsanyi 2013: 1). Of course, in reality, syntactic variation does not, in principle, differ from variation, for example, in accent or dialect lexis. It is possible to find rare syntactic phenomena that are restricted to only one or very few dialects, just like it is possible to find almost universal features represented in all dialect areas (Auer 2004: 75–78; Kortmann 2010: 846).

Another significant reason behind the lack of dialect syntax studies is definitely the methodological difficulties that acquiring syntactic data during fieldwork raises. If a linguist is collecting data with dialectological interviews and questionnaires, it is often difficult or nearly impossible to formulate the right questions that elicit the specific constructions needed from the speaker (Carrilho 2010: 57–58). So while syntax-oriented questionnaires have been compiled for, e.g., Dutch, German and the Nordic dialects (see, e.g., Vangsnes 2007; Berger, Glaser & Seiler 2012), this is uncommon for many languages, including Estonian.

In addition, syntactic variation is typically more subtle than other types of linguistic variation and shows up more as a statistical frequency rather than as a presence or an absence of a variant. When considering dialect isoglosses, it has been shown that dialect syntax phenomena have transitional zones between different variation areas rather than sharp isogloss lines. In the transition zones the frequencies of the morphosyntactic phenomena are lower than in the clearly distinguishable syntactic areas, but the phenomena still occur to some extent. This means that dialect syntax phenomena have a wider areal reach and they are not stereotypically restricted to small areas or individual dialects. In order to study

syntactic variation, the researcher needs significantly more data than other linguistic fields might need: according to some estimates, analysing syntax needs about 40 times as much data as phonetic analysis (Anderwald & Szmrecsanyi 2009: 1136; Kortmann 2010: 842, 844–845, 853).

Starting from the 1990s, dialect syntax began to generate wider interest (Carrilho 2010: 58; Kortmann 2010: 837). This was heavily influenced by the emergence of dialect corpora in many languages, but it was also due to methodological advances that it became possible to study much larger databases than traditional dialectological research had before, and due to dialectology starting to derive methods from sociolinguistics and variation studies (Filppula et al. 2005: vii).

All things considered, in recent years, dialect syntax studies have shifted away from studying isolated syntactic phenomena in isolated dialects toward larger comparative studies of broader regions (if possible, of all dialects of a language) and of complex range of syntactic phenomena, and they have done so in large part thanks to dialect corpora and the emergence of new statistical methods (Kortmann 2010: 842).

3.1.3. Dialectology in Estonia

Differences in Estonian dialects have already been noted since the 17–18th century, as grammarians of the time (e.g., Heinrich Stahl, Anton Thor Helle, Johann Gutschlaff, August Wilhelm Hupel) started showing more interest in dialects in general. They included descriptions of Estonian dialects in their publications and divided them into main dialect groups. 19th century writers (e.g., Johann Heinrich Rosenplänter, Otto Wilhelm Masing, Friedrich Robert Faehlmann, etc.) started publishing dialect texts and discussing the characteristics of and phenomena found in the different dialects (Kask 1984: 29–32; Pajusalu et al. 2020: 35–36; Viikberg 2020: 15).

However, the proper founder of Estonian dialectology is considered to be Ferdinand Johann Wiedemann: during the second half of the 1800s, he collected extensive amounts of dialect data, gave the first comprehensive descriptive overview of the main features of all Estonian dialects and published many studies about the dialects, most of which are still held in high regard today (Kask 1984: 33–34; Pajusalu et al. 2020: 36; Viikberg 2020: 16–17, 24).

Wiedemann was followed by a number of significant researchers in Estonian dialectology, like Jakob Hurt, Mihkel Veske, Heiki Ojansuu, Lauri Kettunen and Jaan Jõgever, among others, who continued to hold lectures and publish studies on dialects, emphasised the importance of dialect research and encouraged the further collection of dialect data, gave instructions on how to phonetically record the data, etc. Kettunen's works from that time are especially noteworthy, as his studies on the Kodavere subdialect were one of the first to thoroughly analyse a subdialect's language and, in addition, novel in their method, combining experimental phonetic analysis with historical-comparative dialectology, juxtaposing the results with other Estonian dialects and closer cognate languages (Kask 1984:

34–38; Pajusalu et al. 2020: 36–37; Viikberg 2020: 17; see also Kettunen 1913a, 1913b).

The 1920s are considered the start of a new period in Estonian dialectology, as the Mother Tongue Society (*Emakeele Selts*) was founded and systematic research of Estonian dialects began. Compiling dialect atlases had gained ground in Europe in the beginning of the 20th century and this practice began in Estonia as well. However, unlike in the rest of Europe, using questionnaires as a method for data collection was less common here. More often, the necessary input for atlases and other studies was acquired by dialect interviews instead (Kask 1984: 40; Pajusalu et al. 2020: 38; Uiboed & Lindström 2014: 14; Lindström & Uiboed 2017: 6; Lindström & Pilvik 2018: 645).

The Mother Tongue Society was one of the first to send a group of researchers to gather data in order to compile exhaustive vocabulary collections, dialect atlases, text selections and phonetic overviews. Lexical information was first collected mostly according to Wiedemann’s dictionary, later by thematic fields, in an attempt not to miss any specific dialect words (Pajusalu et al. 2020: 38).

The main representative in the dialect data collection and the dialect atlas field was Andrus Saareste: his life work amounted to several collections of maps about different features in Estonian dialects (Lindström & Pilvik 2018: 645; see Saareste 1938, 1941, 1955). Another important contribution of his was the four volumes of his thesaurus (Saareste 1958–1968), the aim of which was to help correspondents record even the rarest dialect vocabulary. In addition, he presented his own dialect division in his publications, developed the criteria to distinguish the dialects from each other and determine the borders between (sub)dialects, and gave a detailed account about each dialect’s characteristics. All these have been the foundation of many subsequent dialect studies in Estonia (Pall 1994: 5; Viikberg 2020: 18–19).

An important milestone in Estonian dialectology was the introduction of battery-powered recorders in 1957. Although sound recordings on wax cylinders had been made before, the new recorders brought a new era with them and continuous dialect expeditions to different areas of Estonia were organised every summer from the 1960s up until the 1990s. This enabled the compilation of massive collections of dialect data at the Archives of Estonian Dialects and Kindred Languages at the University of Tartu² and the Archive of Estonian Dialects and Finno-Ugric Languages at the Institute of the Estonian Language³ (Pajusalu et al. 2020: 40; Lindström, Lippus & Tuisk 2019: 328–329).

It is very fortunate that such efforts were made in the last century to collect and preserve dialect data, as, due to the levelling of most dialects in Estonia, it would be impossible to get an accurate representation of them nowadays. Recording and analysing dialect speech continued to be considered important even throughout the Soviet times. As it is, modern researchers have a plethora of dialect overviews, records, voice recordings, etc. at their disposal, allowing researchers to

² <https://murdearhiiv.ut.ee/> (Accessed: January 23, 2023)

³ <http://emsuka.eki.ee/> (Accessed: January 23, 2023)

use this material in many different ways beyond the scope of traditional historical-comparative linguistics (Lindström & Pilvik 2018: 644–645, 648).

During the last decade of the last century, dialectology became increasingly corpus-based. In 1998, the Institute of the Estonian Language and the University of Tartu started compiling the Corpus of Estonian Dialects based on the large dialect collections at both institutions. This corpus mostly reflects the spoken language of speakers born between 1870 and 1890. Conversations and interviews with local speakers have been systematically recorded since the 1950s, then transcribed and morphologically (and at times syntactically) annotated. It is available for use by everyone online⁴. This corpus is also the basis for the research of this dissertation (a further overview of the corpus is given in Section 4.1).

It must be noted that in terms of corpus-based research, Estonian dialects are better off than many other larger languages due to the extensive dialect data collection carried out during the last century. For example, as of November 2022, the Corpus of Estonian Dialects contained 1.24 million morphologically annotated words. In comparison, the Finnish dialect corpus (Lauseopin arkisto⁵) currently contains 1.19 million words, with 887,000 of them annotated (Pajusalu et al. 2020: 40; Uiboed & Lindström 2014: 16; Lindström & Pilvik 2018: 648).

Overall, the 1980–90s can be said to be revolutionary in Estonian dialectology, as new (computerised) methods emerged, opening the way for dialectometry and other research directions (Pajusalu et al. 2020: 40). The first dialectometric studies of Estonian dialects were conducted by Murumets (1982, 1983, 1984) and later continued by Krikmann and Pajusalu (2000) and Pajusalu, Krikmann and Winkler (2009), all focusing mostly on the lexical relations between subdialects. In recent years, dialectometry has been applied to the study of verb constructions (Uiboed 2010a; 2013; Uiboed et al. 2013). In addition to new research directions, new linguistic theories have been increasingly taken into account when studying dialects, with methods from sociolinguistics, spoken language research and other fields being incorporated into dialectology studies (Pajusalu et al. 2020: 40).

The interest in dialectology in Estonia has kept growing over the years and thus there have been a wide range and number of works focusing on many different aspects of dialects. Here I give a short overview of some of the research that has been published, but this list is in no way exhaustive and there are countless more studies that have been conducted in dialectology.

In terms of dialect vocabulary, studies have focused on the distinctive vocabulary of particular dialects, sometimes in comparison with the vocabulary of other dialects (Saareste 1924; Norvik 1985, 2000; Sedrik 1994; Söderman 1996; Koponen 1998; Lindström et al. 2001; Lindström, Velsker, et al. 2009; Pajusalu, Krikmann & Winkler 2009), on different thematic lexis fields like words for birds, plants, fish or colours (Mäger 1967; Vilbaste 1993; Oja 1996; 2001; Kendla 1997, 1999, 2005), on loanwords in Estonian dialects (Vaba 1989a, 1989b, 1997;

⁴ <https://www.murre.ut.ee/mkweb/> (Accessed: January 23, 2023)

⁵ <https://www.kielipankki.fi/aineistot/la-murre/> (Accessed: January 28, 2023)

Juhkam 2000; Must 2000; Björklöf 2012, 2019; Metsmägi, Sedrik & Oja 2014), etc.

Concerning the morphology of dialects, researchers have studied noun morphology (Rosenberg 2000, 2002; Must & Univere 2002; Iva 2007; Metslang & Lindström 2017), verb morphology (Tanning 1965; Pajusalu 1989, 1996, 2005; Toomsalu 1995; Pihelgas 1999; Must & Univere 2002; Iva 2007), plural forms (Neetar 1973), the influences of contact languages on grammar and morphology (Vaba 2010), etc. The South Estonian dialects have been the focus of much research, with studies on the variation of inessive *n-* and *h-* endings (Pajusalu, Velsker & Org 1999; Velsker 2000; Mets 2010, 2011), *nud-* and *tud-* participles (Pihelgas 2000; Mets 2007, 2010, 2013), the quotative (Iva 2002), evidentiality (Metslang & Pajusalu 2002a, 2002b), adjective suffixes (Käsi 2000), etc.

In terms of the phonetic and phonological approach, studies have been done on the vowel system of the dialects (Lindström & Pajusalu 2003), prevocalisation and palatalisation (Org 2003, 2005, 2016; Pajusalu & Teras 2012; Pöld 2019), quantity alternation (Parve 2003; Kalvik 2005; Iva 2010), different speech sounds in the southern dialects (K. Pajusalu 1998; Teras 1998, 2003; Parve 2000; Iva 2003; Lippus 2003), vowels, intonation and other phonetic features in the Insular dialect (Asu 2004, 2005; Niit 2005, 2007; Asu et al. 2012; Asu & Salveste 2012; Türk et al. 2016), pronunciation in the Leivu subdialect (Teras 2007), etc. Viikberg (2020) has compiled a comprehensive overview of the traditional approaches to phonetics and morphology in all the Estonian dialects.

Studies of dialect syntax are addressed in the following section.

3.1.4. Dialect syntax in Estonia

As in the rest of the world, dialect syntax is also a relatively new field in Estonian dialectology. While dialect syntax was somewhat researched before the 21st century, the studies were few (compared to dialect lexis, morphology or phonology), mostly descriptive and dealt with one or a few features in concise dialect areas (Uiboed & Lindström 2014: 19).

To give a few examples, Keem (1961) described archaic object case government in the Nõo subdialect (Tartu dialect). She pointed out that for certain verbs and constructions, the total object is more likely to be in the nominative case, as opposed to the genitive case that is the norm in Standard Estonian and used by younger dialect speakers. Juhkam (1983), in turn, studied the object in the Risti subdialect (Western dialect), where similarly to the Nõo subdialect, the nominative case for objects is preferred over the genitive or partitive. Juhkam proposes that this is due to the influence of Swedish in these dialect areas. Many researchers have examined negation in different Estonian (sub)dialects, e.g., Koit (1963) in the Insular dialect, Sang (1974, 1975) in the Kihnu subdialect and Lindström (1997) in the Võru dialect. Some dialect overviews have also included a section on syntax (see Must 1987: 284–306, 1995: 38–41; Juhkam & Sepp 2000: 27–30, 55–57; Keem & Käsi 2002: 38–53; Lonn & Niit 2002: 53–57; Pajusalu et al. 2020: 112–121). While many of these studies can give a comprehensive overview

of a phenomenon in a particular dialect area, it is difficult based on these to form a complete picture of all Estonian dialects in general (Uiboed & Lindström 2014: 19).

However, there have also been a few large-scale dialect studies encompassing all the dialects in the 20th century. For example, Kask (1984: 251–285) described all the ways in which the quotative mood can be formed, which dialects use which constructions and illustrated the results on a comprehensive map of all Estonian dialects. Among many other syntax studies, Neetar published several articles on the agreement of subject and verb (Neetar 1963, 1964, 1965a, 1965b). She showed that while typically subject-verb agreement exists both for number and person in dialects (similarly to the norm of Standard Estonian), there are many contexts where disagreement can happen, depending on the differences in morphology and the multitude of ways of expressing number and person in different dialects (Neetar 1965b). Nurkse (1937: 31–50) dedicated a chapter of his book on adjective attribute agreement to Estonian dialects, demonstrating, for example, that the disagreement between an allative noun and its adjective modifier is prevalent in the northern, eastern and western dialects, while adjectives typically agree with their noun in southern Estonia. In addition, he showed that syllable economy plays an important part in number agreement, as adjectives that do not have an increase in syllables between singular genitive and plural genitive are overwhelmingly in agreement with their noun. Saareste's (1955) dialect atlas also contains a few syntactic features, like the variation of quotative mood and the distribution of participle versus infinitival constructions.

As the compiling of the Corpus of Estonian Dialects started in 1998, studying dialect syntax in a much more comprehensive manner became possible for many researchers and interest in that branch of dialectology grew. Since then, numerous works have employed the corpus in order to address various syntactic phenomena in the dialects, e.g., Lindström (2001b), Lindström et al. (2006), Kalmus (2009), Lindström, Kalmus, et al. (2009), Lindström, Velsker, et al. (2009), Uiboed (2010a, 2010b), etc.

In 2013, the Estonian Dialect Syntax⁶ project began, with the aim of studying morphosyntactic variation of Estonian dialects using the data from the Corpus of Estonian Dialects. Studies in this project focused on the geographic distribution and variation of frequent morphosyntactic phenomena, using modern statistical methods and putting an emphasis on visualising the data, all in order to provide a more comprehensive picture of Estonian dialect syntax than linguists have ever had before (Uiboed & Lindström 2014: 21).

So far, there have been over 40 articles published and several conferences held as part of this project. Studies have been done on deverbal suffixes (Pilvik 2016, 2017, 2019, 2021), negation (Lindström, Pilvik & Plado 2021; Pilvik, Plado & Lindström 2021), *need*-constructions (Lindström, Uiboed & Vihman 2014; Lindström & Uiboed 2017), verb+verb constructions (Uiboed 2013; Uiboed et al. 2013), the distribution of pre- and postpositions (Ruutma 2016, 2019;

⁶ <https://estdiasyn.ut.ee/project> (Accessed: December 22, 2022)

Ruutma et al. 2016), compound tenses (Lindström et al. 2015, 2019), partitive subjects (Lindström 2017), and clause formation (Plado 2015; Lindström, Pilvik & Plado 2018), among many others. All the papers in this dissertation have also been part of the continuation of the Estonian Dialect Syntax project under the Centre of Excellence in Estonian Studies.

All of these studies have vastly improved our understanding of the complex world of Estonian dialect syntax. To give one example, it has been found that although Estonian dialects are typically divided into southern and northern dialects (which differ the most in phonology, morphology, and lexis), on a (morpho)syntactic level, an imaginary line can instead be drawn between eastern and western dialects, with the Coastal and Mulgi dialects fitting in with either group depending on the studied phenomenon (see Lindström, Kalmus, et al. 2009; Uibo 2013; Uibo et al. 2013; Lindström, Uibo & Vihman 2014; Lindström et al. 2015; Ruutma et al. 2016; Lindström & Uibo 2017; Lindström, Pilvik & Plado 2018). This result is also seen in P3.

3.2. Variation

Over the past few decades, the study of language variation (and change) has become one of the most productive and successful fields of linguistic research (Krug, Schlüter & Rosenbach 2013: 1). As this dissertation also studies syntactic variation of pronouns in Estonian dialects and employs the methods of variation analysis, in this section I describe the main principles of studying variation in language and give a brief overview of the history of the field.

Variation analysis begins with the statement that there are two or more ways of saying the same thing, i.e., there exists a linguistic variable whose different variants cannot be clearly distinguished from each other in either meaning or in function. However, the different forms of a variable can vary in terms of social, stylistic and/or linguistic parameters (Labov 2004: 7; Feagin 2008: 23; Walker 2013: 440–441). This variation can be found on all linguistic levels, in all languages and their varieties, in all registers and dialects, in every person's speech, sometimes even in the same discourse and in the same sentence. Simply put, language variation exists everywhere (Tagliamonte 2013: 383).

The concept of a (socio)linguistic variable comes from the early works of William Labov (1963, see also 1966), where he made a connection between the phonetic shift in certain diphthongs and the speakers' (subconscious) attitudes towards their place of residence. His studies were one of the first in the field where the central focus was on the variable itself; before Labov, linguistic variation was mostly considered to be unstructured, unimportant and random. It was not only thought to be not worth researching, but linguists of the time often even used invented or "cleaned-up" data in order to avoid the variation in naturally occurring data (Milroy & Milroy 1997: 47–48). Of course, by now the study of variation has shown that variation on any level of linguistics is not random or

free, but is most likely influenced by social and/or linguistic factors (Chambers & Trudgill 1998: 49–50; Bayley 2008: 117, 122).

This is exactly what matters the most for a variationist linguist: not that a linguistic variation is happening, but that the variation is correlated with other social and/or linguistic factors (Chambers & Trudgill 1998: 70). Based on the choice of one variant over the other we can interpret a lot about the processing of the variable, about the constraints on using that particular variable and producing it in real time, about the characteristics of the speaker, about its social value and also about the general tendencies in language. It is interesting that people are often unaware that they are using particular linguistic variables at all. For many variables it has been shown that a majority of the speakers use both (or more) of the studied variants, but the proportions of the use of one or the other are correlated with social variables (Chambers & Trudgill 1998: 50; Tagliamonte 2012: 6; Krug, Schlüter & Rosenbach 2013: 1). In addition, morphosyntactic variables are considered to mark social differences more dramatically than phonetic ones (Tagliamonte 2012: 206).

For obvious reasons, studying variation in phonetics or morphology is more straightforward, as it is fairly simple to determine the different variants. For these fields, a form-based approach is often used: the researcher determines a set of variables that occur in the same context or have an identical meaning. In syntax, lexis and discourse, however, variation is often more difficult to define. The equivalence and synonymy of two variants is much vaguer and it is often impossible to confidently know whether the two forms are really saying the same thing or whether each variant is distinguished from others by miniscule semantic nuances. For syntactic or textual analysis, a function-based approach is more suitable: first, one defines a particular linguistic function, and then, finds all the possible forms that fill that function (Cameron & Schwenker 2013: 468; Walker 2013: 440–443).

Recent decades have brought a quantitative turn to linguistics and, therefore, to variation research, and the amount of corpus-based and probabilistic grammar studies has exponentially increased. While probabilistic grammar analysis has been around for a while, it is only in recent years that it has begun to be explored in a more systematic way. Probabilistic variation implies that the speakers' implicit knowledge of language is always to some extent probabilistic, as the likelihood of using a linguistic variant in a certain context tends to correspond with the speakers' intuition about the acceptability of the variants. In a sense, it says that variation is dependent on many constraints, which sometimes overlap and compete with each other and which all influence the speakers' linguistic choices in probabilistic ways, leading, in turn, to the statistical variability in the distribution of different forms and variants (Szmrecsanyi 2017: 694; Grafmiller et al. 2018: 1–4; see also Bod, Hay & Jannedy 2003; Bresnan et al. 2007).

A probabilistic grammar approach to variation is inherently usage-based, as, unlike sociolinguistic variation studies, which focus on demographic factors like age, gender, education, etc., it also associates the variation patterns with language-internal factors, like lexical items or forms, linguistic context, abstract constraints

(like animacy in this dissertation), register, speakers' cognitive abilities, etc., in order to figure out what all these factors can reveal about the speakers' implicit linguistic knowledge (Szmrecsanyi 2017: 688, 693; Grafmiller et al. 2018: 3).

Corpus-based probabilistic grammar research differs from traditional socio-linguistic variation research in many ways. For example, its focus is shifted from phonology to lexical and grammatical variation. In addition, while typical variationist linguistics has not been interested in written material and has preferred analysing spoken data, corpus-based variation studies have started to include other styles and registers, already showing that the variation patterns of vernacular speech can be generalised to written text as well. Furthermore, corpus-based studies can include more than one variable (although many works are still based on a single feature) in order to get more reliable results through investigating how the different variables are interacting with each other. Corpus-based probabilistic variation has been the driving force behind the development of new statistical methods and techniques, broadening the analysis possibilities from the traditional binary logistic regression of variationist sociolinguistics (Schneider 2008: 67; Szmrecsanyi 2017: 689–690, 695; see also Guy 2005).

In Estonia, the study of variation began in the 1990s, led by Karl Pajusalu, and at first, as in the rest of the world, the emphasis was more on sociolinguistic variation (Lindström & Pilvik 2018: 647). Examples of sociolinguistic studies include 1st person pronoun ellipsis in all Estonian dialects (Lindström, Kalmus, et al. 2009), the variation of word-final [e] in the Karksi subdialect (K. Pajusalu 1996), and the variation of the inessive case ending (Pajusalu, Velsker & Org 1999; Velsker 2000) and vernacular speech variation and changes (Mets 2010) in the Võru dialect, among others.

Later dialect variation studies, many of which have been based on corpus material, have focused instead on explaining variation through language-internal factors and probabilistic linguistics. This is partly also due to the nature of the dialect corpus, the speakers of which have all been chosen on a similar basis: older people, often women, who have not moved much around during their life and have not received much education (see more in Section 4.1). This means that there is not enough variability in the background of the speakers to conduct a reliable sociolinguistic analysis. As examples, dialect studies have been done on the variation of synthetic adessive and analytic adpositional constructions (Klavan, Pilvik & Uibo 2015), *need*-constructions (Lindström & Uibo 2017), naming constructions (Lindström, Pilvik & Plado 2018), negation in the Võru and Seto dialects (Lindström, Pilvik & Plado 2021; Pilvik, Plado & Lindström 2021), etc. All the papers in this dissertation also contribute to the probabilistic study of variation in Estonian dialects.

4. DATA AND METHODS

4.1. Data

The articles in this dissertation used two corpora as their data source: the Corpus of Estonian Dialects (CED) and the Phonetic Corpus of Estonian Spontaneous Speech (PCESS). As this dissertation is first and foremost a study of dialectology, the aim of which has been to examine the syntactic variation of pronouns in Estonian dialects, the CED was the most important source material: it is used in all the publications of the dissertation and in the additional combined analysis of all three studied pronouns (see Sections 5.2 and 5.3).

This section gives a thorough overview of the CED and the process of obtaining data from the corpus. In P3, the PCESS has been included in order to compare the variation in the dialects with the variation in contemporary speech. An overview of the PCESS and its numerical parameters during the data analysis can be found in P3.

The Corpus of Estonian Dialects⁷ is an electronic collection of authentic spoken texts, in which all Estonian dialects are represented. The language recorded in these texts is as archaic as possible, with most of it having been recorded in the 1960s and 1970s, and the earliest recordings dating back to 1938. All of the original recordings are stored at the Institute of the Estonian Language and at the University of Tartu Archives of Estonian Dialects and Kindred Languages. Since 1998, the CED has been developed by the Institute of Estonian and General Linguistics at the University of Tartu in cooperation with the Institute of the Estonian Language (Lindström, Lippus & Tuisk 2019: 337).

The CED consists of the following (Lindström, Lippus & Tuisk 2019: 337):

1. dialect recordings;
2. phonetically transcribed texts;
3. texts in simplified transcriptions;
4. morphologically annotated texts;
5. a separate database for information about the informants, recordings and transcriptions.

The recordings are traditional dialect recordings, where the informant is usually interviewed at their home. Informants are typically older people (about 70–90 years old) and mostly women. The topics covered in the recordings are similar throughout the corpus, concerning the informant's personal life, their lifestyle, past events and working methods (Lindström, Lippus & Tuisk 2019: 338). In a sense, the CED can be considered to be a corpus of a historical language, as it does not

⁷ <https://datadoi.ee/handle/33/492> (Accessed: January 26, 2023)

represent modern dialects, which in many areas of Estonia have unfortunately disappeared or been severely levelled due to urbanisation and the influence of Standard Estonian.

There are several different ways that the Estonian dialects can be and have been divided (a thorough overview of dialect division in the past and the present can be found in Antso (2018) and Pajusalu et al. (2020: 48–54)). In the CED, dialects have been divided into ten areas: the Mid, Western, Insular and Eastern dialects make up the North Estonian dialect group, the Coastal and Northeastern (or Alutaguse) dialects make up the Northeastern-Coastal dialect group, and the Tartu, Võru, Mulgi and Seto dialects make up the South Estonian dialect group. This division differs from the traditional dialect grouping, where the Northeastern and Coastal dialects have been regarded as one dialect and the Seto dialect has been considered a subdialect of Võru (see Kask 1984). The dialect areas of the CED are shown in Figure 1.

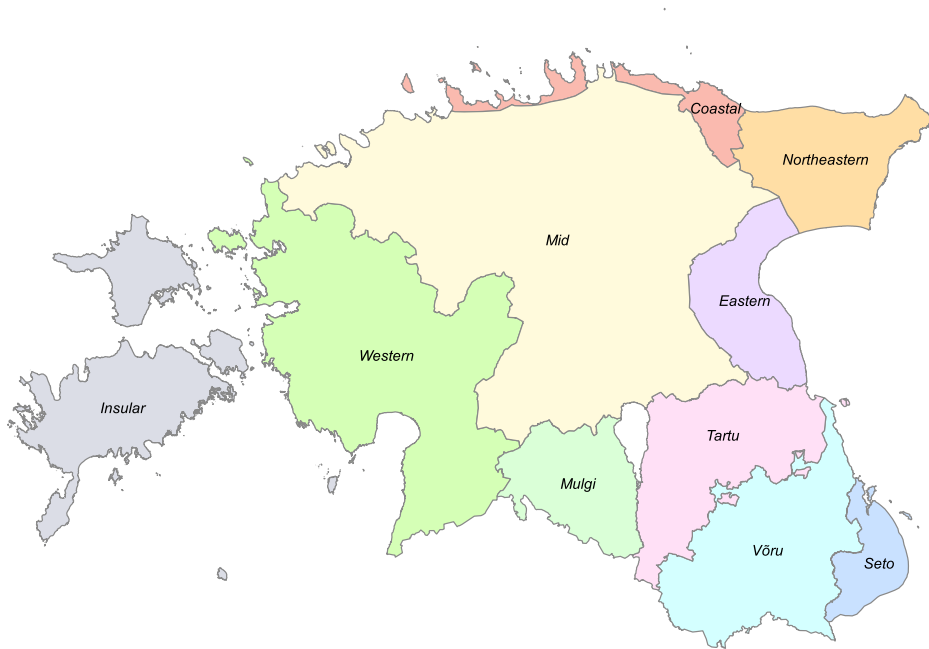


Figure 1. The dialect division of the CED

All the dialects in the corpus have been divided into subdialects (the borders of which are based on former administrative units – parishes). The subdialects represented in the corpus data used in this dissertation are shown in Figure 2. It is clear that the representation of subdialects differs between dialects, e.g., the data contain all of the subdialects of the Northeastern and Võru dialects, but only about half of the subdialects of the Mid dialect are represented (mostly due to it being the largest dialect area). However, although it should be noted that the data

Table 8. The number of morphologically annotated tokens, texts, informants and sub-dialects in the CED in January 2019

| <i>Dialect</i> | <i>Total tokens</i> | <i>Texts</i> | <i>Informants (m/f)</i> | <i>Subdialects</i> |
|----------------|---------------------|--------------|-------------------------|--------------------|
| Western | 251,031 | 82 | 63 (21/42) | 14 |
| Mid | 246,167 | 83 | 72 (24/48) | 19 |
| Insular | 202,325 | 69 | 55 (10/45) | 15 |
| Võru | 111,503 | 41 | 34 (8/26) | 8 |
| Coastal | 97,152 | 29 | 24 (10/14) | 4 |
| Tartu | 80,343 | 28 | 24 (6/18) | 7 |
| Seto | 68,414 | 19 | 20 (4/16) | 3 |
| Mulgi | 63,759 | 20 | 18 (6/12) | 4 |
| Northeastern | 60,037 | 19 | 14 (7/7) | 4 |
| Eastern | 48,353 | 23 | 20 (8/12) | 4 |
| Σ | 1,229,084 | 413 | 344 (104/240) | 82 |

In this dissertation I have used the (manually) morphologically annotated dialect texts, which are in XML-format and where each token in the texts has been annotated with its lemma, word class and word form. In all the examples in this dissertation, all transcription markings (palatalisation, stress, quantity degrees, compound word borders, etc.) have been removed, as they are not relevant to the studies. However, the orthography of the examples is presented exactly as it is in the corpus.

In order to acquire the necessary info from the XML-files, I have used a Python script by Kristel Uibo⁸: inserting the lemma and context window, the script extracts all the cases of that lemma from all the texts and combines them into an Excel file. Every keyword in that Excel file is accompanied by the preceding and following context (up to 20 words), case marking and information about the informant (dialect, subdialect, village, age, gender, date of birth and place of birth). In this dissertation I have mostly compared the language of the dialects and not the subdialects, as typically only a few speakers represent a subdialect, which is not enough for a comprehensive analysis.

4.2. Methods

This section gives an overview of all the statistical methods used in this dissertation: conditional inference trees and random forests (P1, P2, P3), mixed-effects logistic regression (P3) and multiple correspondence analysis (P2). In addition to the published articles, conditional inference trees and multiple correspondence

⁸ <https://github.com/kristel-/RuRaKe/blob/master/Scripts/CedConcordancer.py> (Accessed: February 3, 2023)

analysis are employed in Sections 5.2 and 5.3 of this overview article. These methods are briefly described in the following sections.

Each of these methods analyses the data in their own unique way, yet can be complimentary to each other when used together. Logistic regression has been used in quantitative variation studies for a while, but it is a more recent approach to apply mixed-effects models, random forests and conditional inference trees to the variation data. Furthermore, studies like Tagliamonte and Baayen (2012) and Baayen et al. (2013) have shown that it is beneficial to use all of these complementary techniques together. Mixed-effects models provide a way to account for random-effect factors (e.g., characteristics of the speakers in the data). Random forests, in turn, provide information about the importance of predictors. Finally, conditional inference trees and correspondence analysis visualise how different predictors operate in association with each other.

Studying (dialect) syntax poses a problem where, on one hand, it is necessary to have a large amount of natural language recordings (as it is often difficult to reproduce syntactic phenomena in a controlled environment), but, on the other hand, this kind of data collection can result in an unpredictably unbalanced dataset. That means that the phenomenon of interest can be represented many times in one dialect area or construction and hardly ever in another due to arbitrary and uncontrollable factors during data collection, rather than due to the actual distribution of the phenomenon. However, as conditional inference trees, random forests and multiple correspondence analysis place no particular requirements upon the data, they are highly suitable to use as statistical methods in case of an unbalanced dataset with mostly categorical variables.

All these methods have also been successfully used in other studies of Estonian, Estonian dialects and (dialect) syntax (see, e.g., Uihoaed 2013; Klavan, Pilvik & Uihoaed 2015; Reile 2015; Ruutma et al. 2016; Lindström & Uihoaed 2017; Taremaa 2017; Lindström, Pilvik & Plado 2018; Proos 2020; Tragel & Klavan 2021; Hint et al. 2021; Lindström, Pilvik & Plado 2021; Pilvik, Plado & Lindström 2021).

All the calculations were performed using the statistical software R (R Core Team 2018). The conditional inference trees and random forests were computed using the functions *ctree()* and *cforest()* from the package *party* (Hothorn, Hornik & Zeileis 2006). The binomial mixed-effects logistic regression was computed using the function *glmer()* from the package *lme4* (Bates et al. 2015). The multiple correspondence analysis was computed using the function *mjca()* from the *ca* package (Nenadic & Greenacre 2007).

4.2.1. Conditional inference trees and random forests

Conditional inference trees and random forests are methods based on binary recursive partitioning. At each stage of the analysis, the tree model's algorithm tests the association between the independent variables and the given response variable. The variable most strongly associated with the response variable is the one used to split the data into two subsets. The two subsets are then tested again for

the association between the variables and if possible, split into two. This kind of partitioning continues until no variable is associated with the response at a level of statistical significance. The result of this process is presented as a tree structure with binary splits (Hothorn, Hornik & Zeileis 2006; Strobl, Malley & Tutz 2009).

For random forests, the model outputs a measure of importance for each variable, averaged over many conditional inference trees. These (permutation) variable importance measures, in turn, reflect the impact value each variable has on the response. The conditional importance scheme has been used for the computation of these measures in order to take into account the possible correlations between the variables and to help identify the truly relevant predictor variables (see Strobl et al. 2008). The goal of both conditional inference trees and random forests is to predict the chances of the dependent variable occurring in a given context, specified by the independent variables (Breiman 2001).

4.2.2. Binomial mixed-effects logistic regression

Binomial mixed-effect logistic regression is a method suitable for explaining and predicting speakers' choice between two variants on the basis of geographic, social, syntactic, pragmatic, and other factors. A mixed-effects model includes fixed effects, which are known as normal explanatory variables and the effect of which is measured, and random effects, which are sampled randomly from the population (e.g., individual speakers, lexical stimuli, individual lemmas, etc.) and that are useful in filtering out “noise” from the model. Logistic regression takes a pre-determined configuration of predictors as a reference and compares the change in effect on the response for each predictor value, while keeping the other predictors constant (Fielding & Goldstein 2006; Tagliamonte & Baayen 2012; see also Baayen, Davidson & Bates 2008).

Logistic regression models are more particular about the distribution of the datasets and set certain requirements for the data. One of those requirements concerns observed values: if a variable's value occurs fewer times than 5–10 (the exact threshold is under debate by researchers), the model's estimation can be wrong (Walker & Smith 2020). This kind of sparseness of data, however, is a common problem when studying dialects and especially dialect syntax. Thus logistic regression is only used in P3, where combining the data from the Corpus of Estonian Dialects and the Phonetic Corpus of Estonian Spontaneous Speech fits the necessary requirements that the dialect data on their own could not.

4.2.3. Multiple correspondence analysis

Correspondence analysis (CA) is an exploratory technique designed specifically for the analysis of categorical variables. CA takes the frequency of co-occurring features and converts them into distances, which are then plotted on a two- or three-dimensional graph to visualise how the variable values are associated with each other (Glynn 2014: 445). Multiple correspondence analysis (MCA) is an

extension of CA and has the ability of analysing more than two variables simultaneously (Hill & Lewicki 2006: 136).

For most data, the combination of the first two dimensions of MCA offers the most accurate and easily interpretable visualisation of how the variables and their values are associated with each other (Glynn 2014: 447). The further a value on a plot is from the origin (the point where the x-axis and y-axis intersect), the more discriminating it is. Inversely, the closer a value is to the origin, the less discriminating it is, but only in the context of the chosen variables. This means that a variable or a value might still contribute to the studied variation, but not in the visualised dimensions (Alberti 2015: 27; see also Greenacre 2017: 97–104).

To analyse the relationship between one variable's value with another variable's value, one should look at the angle connecting the two values via the origin: the smaller the angle, the stronger the positive association probably is. If the angle is 90 degrees, the values are most likely not associated at all and if the angle is 180 degrees, the values are probably negatively associated with each other (Yelland 2010: 102; Glynn 2014: 455–459; Greenacre 2017: 86).

However, MCA does not show whether the associations between the variable values are significant or relevant at all, since the primary purpose of this technique is just to produce a simplified representation of the data. Therefore, one must check all conclusions made with MCA using raw data and other statistical methods (Greenacre 1984: 10; Hill & Lewicki 2006: 134; Glynn 2014: 444).

5. RESULTS AND DISCUSSION

In this chapter, I first summarise the results of the published articles of this dissertation in Section 5.1. As the animacy reference of the pronoun *mis* was not addressed in any of the publications, Section 5.2 contains a brief analysis of how *mis* behaves in terms of referring to animate and inanimate entities and which variables significantly affect its use. In Section 5.3, I combine the datasets of *kes*, *mis* and *keegi* and conduct a comprehensive analysis of their animacy reference, comparing the use of the three pronouns in Estonian dialects.

5.1. Overview of the publications

P1. The pronoun *kes* ‘who’ and its referent’s animacy in Estonian dialects

The first article of the dissertation concerns the use of the interrogative-relative pronoun *kes* ‘who’ in Estonian dialects. There is a clear distinction between the pronouns *kes* and *mis* ‘what’ in Standard Estonian: *kes* refers to animate entities and *mis* refers to inanimate entities. However, in Estonian dialects it is not always so and *kes* can sometimes refer to inanimate entities and *mis* to animate entities. The aim of this article was to determine which Estonian dialects typically use *kes* to refer to inanimate entities and which variables most affect this use. The data used came from the Corpus of Estonian Dialects and in order to ascertain the relevance and effect of the chosen variables, the conditional inference tree method was applied to the data.

The analysis showed that *kes* was most likely used to refer to inanimate entities in the entire North Estonian dialect group (except on the island of Saaremaa) and in the Coastal dialect. Southern dialects could be excluded from the typical area of this variation, but elsewhere in Estonia there were many isolated subdialect groups where *kes* was frequently used to refer to inanimates.

Out of all construction types, relative clauses showed the highest percentage of non-selectivity, while there were hardly any instances of *kes* referring to inanimate entities in other construction types. This is most likely due to the grammaticalisation of the interrogative marker into a relative marker: during the grammaticalisation process, its semantics have bleached and it has lost some of its semantic distinctions, one of them being the distinction in animacy. When referring to inanimates, the case of the pronoun was mostly elative or comitative, at times also genitive, while the use of other cases was rare.

In the case of inanimate entities, only postnominal relative clauses were used, while relative clauses that modified a noun referring to an animate entity were both pre- and postnominal. This could, once again, be due to the grammaticalisation of the interrogative marker into a relative marker, the purpose of which in postnominal relative clauses (where the pronoun immediately or closely follows the referenced entity) is not to add any semantic value, but to function only as a

connective relative marker, and therefore its clear distinction in animacy is not as relevant as it would be in prenominal relative clauses.

While the studied variables had complex interactions with each other, the choice to use the pronoun *kes* to refer to inanimate referents was mainly associated with the case of the pronoun (and the relations those cases represented), with dialect and construction type also being significant in determining the speakers' choice in referring to animate or inanimate entities. This all shows that although the pronoun *kes* is still mainly used to refer to animate entities in the dialects, there is significant variation in its use and animacy as a category is not as semantically significant in the dialects as it is in Standard Estonian.

P2. The use of the indefinite pronoun *keegi* 'someone' in Estonian dialects

Furthering the analysis of P1, I continued the study of animacy reference with the indefinite pronoun *keegi* 'someone, no one' in P2. Similarly to *kes* and *mis*, the indefinite pronouns *keegi* and *miski* 'something, nothing' are also used to refer to animate and inanimate entities, respectively, in Standard Estonian. Yet in Estonian dialects, *keegi* is frequently used to refer to inanimate objects as well. In addition to determining which variables affect the choice to refer to either animate or inanimate entities with *keegi*, this paper aimed to describe the functions in which *keegi* can be used in Estonian dialects. The research was based on the data in the Corpus of Estonian Dialects, with conditional inference trees, random forests and multiple correspondence analysis as the chosen statistical methods.

In Estonian dialects, the pronoun *keegi* was used in the following functions: as a nominative and a partitive subject, an object, an adverbial, a genitive and a postnominal attribute, a determiner, a negative polarity item and a generalising alternative. Almost half of all the uses were subjects, but objects, adverbials and negative polarity items were also very frequent.

keegi was most often used to refer to inanimate entities in the Western, Mid and Eastern dialects, where over half of the uses of *keegi* referred to inanimates. At the same time, it did not seem to be possible to use *keegi* to refer to inanimate entities in the Võru and Seto dialects. In terms of animacy reference, the distribution of the dialects was similar to the results of P1, showing us that such developments are probably not coincidental, but rather in those particular areas the animate-inanimate distinction has for some reason started to fade.

The indefinite pronoun was most often used to refer to inanimate entities when *keegi* was an object, a partitive subject or a negative polarity item, when it was in the partitive case and positioned at the end of a negative clause. When *keegi* referred to an animate being, it was most likely a nominative subject at the beginning of an affirmative clause. As shown in P1, the distinction in animacy for the pronoun *kes* was most prevalent when *kes* was used as a relative pronoun (as opposed to an interrogative pronoun), i.e., in the most grammaticalised position. A similar connection can be made for *keegi*, as using it in the most grammaticalised function – as a negative polarity item – has also increased its non-selectivity in animacy.

The results showed how tightly indefinite pronouns and partitive case marking are interrelated in the scope of negation, as well as how the animate-inanimate distinction has become irrelevant in this specific context. The inanimate use of *keegi* in Estonian dialects seems to be following the path of grammaticalisation, as the loss of semantic distinctions can be an early stage in the grammaticalisation process. This can, in turn, potentially result in the pronoun developing into a negation word or a polarity item. However, this grammaticalisation does not seem to affect the animacy distinction very much in syntactic positions which are outside of the scope of negation and for which differentiating between animate and inanimate referents is still relevant for understanding the content of the clause, such as for nominative (canonical) subjects or attributes.

P3. Object case variation of the pronoun *mis* ‘what’ in spontaneous spoken Estonian and Estonian dialects

P3 examined the object case variation of the pronoun *mis* ‘what’ in the position where partitive marking of the object is expected. While nouns in partial object position are characteristically in the partitive, the pronoun *mis* acts irregularly in this position: partitive *mida* can be replaced with nominative *mis* (the case typically used for total objects). This variation occurs in both spoken and written Estonian as well as in Estonian dialects and is characteristic only of the pronoun *mis*.

The aim of this paper was to determine which variables affect this object case variation most significantly, how the variation differs between contemporary speech and antiquated dialects and what might possibly have caused or influenced this variation. The analysis was based on the data from the Corpus of Estonian Dialects and the Phonetic Corpus of Estonian Spontaneous Speech.

While the two corpora differ in their range of variation of *mis* and *mida* in partial object position (in the contemporary data, the division of *mis* and *mida* is more or less equal, but in the dialects, *mida* occurs in less than 10% of the sentences), there are several morphosyntactic and geographical variables that contribute to the variation in both corpora. The conducted analyses showed that verb type, construction type, linguistic context, polarity and dialect are the most significant variables affecting this particular variation: the use of partitive *mida* was more likely to occur in the case of a negative partitive verb and in the easternmost dialects, while its use was less likely when followed by a short (personal) pronoun and with (rhetorical) questions. In this study, the variation of *mis* and *mida* was examined from a morphosyntactic perspective; however, it is likely that some prosodic variables, like intonation and sentence stress, might also contribute to this variation. A further prosodic study, therefore, would be necessary for a comprehensive understanding of this object case alternation.

I proposed that there could be (at least) three possible motivations for the use of *mis* instead of *mida* in partial object position – language contact, high usage frequency and language standardisation – and it is likely that all three may have had an influence of this variation in association with each other. Firstly, while it

is difficult to determine whether any local contacts have been responsible for the increase in prevalence of nominative *mis*, the lack of general Indo-European or specifically Germanic contacts in the eastern dialects (which have retained the use of partitive *mida*) could explain why the Coastal, Northeastern and Seto dialects differ from the other dialects. Secondly, frequent (fixed) constructions showed a higher percentage of nominative *mis* than low-frequency constructions, indicating that high usage frequency could be the motivation behind the choice to use the shorter nominative in partial object position. Finally, the distinct differences in the range of variation in contemporary speech and antiquated dialects could be explained with the effects of language standardisation, which has surely had an influence on the contemporary speech of Estonia, but less so on the dialects of primarily oral tradition.

5.2. The animacy reference of *mis*

The articles included in this dissertation focused on the animacy reference of the interrogative-relative pronoun *kes* and the indefinite pronoun *keegi*, as well as the object case variation of the interrogative-relative pronoun *mis*. However, another topic must be discussed in order to obtain a full picture of the animacy-related variation of the interrogative-relative and indefinite pronouns: the animacy reference of the pronoun *mis*. In a similar fashion to *kes*, while *mis* is used only to refer to inanimate entities in Standard Estonian, in some Estonian dialects a such restriction does not exist and *mis* can refer to both animate and inanimate referents⁹.

In this section, I intend to give a brief overview of the way *mis* functions in terms of its animacy reference and which variables affect the speaker's choice of referring to animate or inanimate entities with this pronoun.

As with *kes* and *keegi*, the data come from the Corpus of Estonian Dialects (see Section 4.1). Each instance of the pronoun *mis* in the dataset has been collected from the morphologically annotated corpus texts, accompanied by the preceding and following context (up to 20 words), case marking and metadata about the informants.

In total, there were 7,558 instances of *mis* in the corpus, 6,111 of which were included in the dataset (excluding incomplete sentences and repetitions). In addition to the variables of case and dialect, which were acquired from the corpus, I have manually added the following variables to each instance: the animacy of the referent, construction type and the polarity of the clause.

Similarly to P1 and P2 of this dissertation, the variable of animacy has two levels: animate and inanimate. As mentioned in Section 2.4, this distinction follows Fowler (1977: 16–17) and categorises all humans and animals as animate,

⁹ *keegi* also has an inanimate counterpart *miski* 'something', but based on the data in the Corpus of Estonian Dialects, it seems that *miski* has not lost any of its animacy distinction in the dialects and is only used to refer to inanimate referents. Therefore, it is not analysed in this dissertation.

and everything else as inanimate. Although animacy in language should be regarded as a continuum rather than a binary variable, the decision to use a binary approach is partly due to the nature of the dialect texts, which rarely have any borderline cases of animacy, and partly due to the nature of the studied pronouns, which often do not have a counterpart on the extended animacy hierarchy.

The construction type variable has seven levels, some of which are very common constructions like relative and interrogative clauses, fixed constructions, and rhetorical and indirect questions, while others are rather characteristic of spoken language, like relative clauses that lack a main clause that they are supposed to be modifying (these typically either introduce a new topic or lack a main clause simply because the speaker's train of thought changed mid-sentence). The seventh type is other constructions that do not fit under any of the previous types. As the clauses containing the pronoun *mis* were divided into construction types exactly as was done for the pronoun *kes*, a more detailed overview of this categorisation can be found in P1.

The polarity of the clause can be either affirmative or negative.

Out of the 6,111 *mis* pronouns in the dataset, 5,688 (93%) refer to inanimate entities and 431 (7%) refer to animate entities. The frequency of the animate and inanimate referents in all the dialects is presented in Table 9.

Table 9. The frequency of animate and inanimate referents of the pronoun *mis* by dialect

| <i>Dialect</i> | <i>Inanimate</i> | <i>Animate</i> | Σ |
|----------------|------------------|-------------------|--------------|
| Northeastern | 216 | 22 (9.2%) | 238 |
| Mid | 1,127 | 111 (9.0%) | 1,238 |
| Western | 1,094 | 98 (8.2%) | 1,192 |
| Mulgi | 343 | 28 (7.6%) | 371 |
| Insular | 1,125 | 82 (6.8%) | 1,207 |
| Tartu | 400 | 25 (5.9%) | 425 |
| Võru | 476 | 27 (5.4%) | 503 |
| Seto | 250 | 11 (4.2%) | 261 |
| Eastern | 258 | 11 (4.1%) | 269 |
| Coastal | 391 | 16 (3.9%) | 407 |
| Σ | 5,680 | 431 (7.0%) | 6,111 |

As we can see from Table 9, it is possible to refer to both animate and inanimate referents with *mis* in all Estonian dialects. Animate reference is most common in the Northeastern (example 6), Mid, Western and Mulgi dialects, while it is rarer in the Seto, Eastern and Coastal dialects. Nevertheless, it is clear that *mis* is predominantly used to refer to inanimate referents, as in Standard Estonian, and even in those dialects where reference to animate beings is more common, it does not occur more than 10% of the time.

- (6) Northeastern (Lüg)
sie õte mitü saunasse vie suoja pani
 that sister what:PRT sauna:ILL water:GEN warm:ILL put:PST:3SG
 ‘The sister who set the water to warm in the sauna.’

Of all construction types, *mis* appears most often in relative clauses (example 7), rhetorical questions and indirect questions (see Table 10). *mis* is most likely to refer to animate entities in relative clauses, with 13.3% of the relative pronouns referring to animate beings. In interrogative clauses, fixed constructions and other constructions, the speakers do not use *mis* to refer to animate entities. Very similar results were obtained in P1 for the use of the pronoun *kes*.

- (7) Insular (Jäm)
kui ma laps oli siis oli söuksed
 when I child be.PST:1SG then be:PST:3PL this_kind:PL
missel olid pisikkused kõrvad
 what:ADE be:PST:3PL tiny:PL ear:PL
 ‘When I was a child, there were the kind [of sheep] that had tiny ears.’

Table 10. The frequency of animate and inanimate referents of the pronoun *mis* by construction type

| <i>Construction type</i> | <i>Inanimate</i> | <i>Animate</i> | Σ |
|-------------------------------|------------------|-------------------|--------------|
| relative clause | 2,430 | 373 (13.3%) | 2,803 |
| rel. clause w/o a main clause | 147 | 5 (3.3%) | 152 |
| rhetorical question | 1,753 | 40 (2.2%) | 1,793 |
| indirect question | 1,044 | 13 (1.2%) | 1,057 |
| interrogative clause | 184 | 0 (0.0%) | 184 |
| fixed construction | 50 | 0 (0.0%) | 50 |
| other | 72 | 0 (0.0%) | 72 |
| Σ | 5,680 | 431 (7.0%) | 6,111 |

Heine & Kuteva (2006: 227) have proposed that as interrogative markers developed through four possible stages of grammaticalisation (from being just an interrogative marker to introducing headed relative clauses), their distinctions in gender, animacy, number and case also decreased down through the stages (see Section 2.4). This corresponds well with these results, as in relative clauses, *mis* is the least systematically distinguished by animacy and its non-selectivity has increased, while in interrogative clauses the distinction in animacy is clear-cut and *mis* can only refer to animate entities.

Of the 14 Estonian cases, the data include nine: the nominative, genitive, partitive, inessive, elative, allative, adessive, translative and comitative. Almost all of the instances of *mis* are in the nominative (92.8%), with the partitive (4.6%) and comitative (1.5%) following at much smaller percentages. Table 11 shows that the case most likely to refer to an animate entity is the adessive; however, as there

are only eight instances of adessive *mis* in the data, no solid conclusions can be made about this. It is more reasonable to conclude that the nominative (example 8) is the case that most probably refers to an animate entity (7.3%), followed by the partitive (4.3%).

- (8) Western (Tõs)
- | | | | | | |
|---------------------|-----------|--------------|------------|-------------|--------------|
| <i>nüüd</i> | <i>oo</i> | <i>see</i> | <i>eit</i> | <i>miss</i> | <i>sinna</i> |
| now | be.3SG | this | old_lady | what | there:ILL |
| <i>minijas</i> | | <i>tuli</i> | <i>koa</i> | <i>seäl</i> | <i>üksi</i> |
| daughter-in-law:TRL | | come:PST:3SG | also | there | alone |
- ‘Now that old lady, who went there as a daughter-in-law, is also alone there.’

Table 11. The frequency of animate and inanimate referents of the pronoun *mis* by case

| <i>Case</i> | <i>Inanimate</i> | <i>Animate</i> | Σ |
|-------------|------------------|-------------------|--------------|
| adessive | 5 | 3 (37.5%) | 8 |
| nominative | 5,255 | 415 (7.3%) | 5,670 |
| partitive | 266 | 12 (4.3%) | 278 |
| comitative | 92 | 1 (1.1%) | 93 |
| allative | 1 | 0 (0.0%) | 1 |
| elative | 32 | 0 (0.0%) | 32 |
| genitive | 17 | 0 (0.0%) | 17 |
| inessive | 1 | 0 (0.0%) | 1 |
| translative | 11 | 0 (0.0%) | 11 |
| Σ | 5,680 | 431 (7.0%) | 6,111 |

In the data, 98.1% of all the clauses containing *mis* are affirmative. When looking at the division of affirmative and negative clauses in terms of animacy, we can see from Table 12 that while it is possible for *mis* to refer to both animate and inanimate entities in clauses with both polarities, neither affirmative nor negative clauses stand out as more likely to refer to animate beings. So it is most probable that polarity does not affect the speaker’s choice in referring to animate or inanimate entities with the pronoun *mis*.

Table 12. The frequency of animate and inanimate referents of the pronoun *mis* by polarity

| <i>Polarity</i> | <i>Inanimate</i> | <i>Animate</i> | Σ |
|-----------------|------------------|-------------------|--------------|
| negative | 108 | 10 (8.5%) | 115 |
| affirmative | 5,572 | 421 (7.0%) | 5,984 |
| Σ | 5,680 | 431 (7.0%) | 6,111 |

Figure 3 shows the conditional inference tree for the animacy of the referent of the pronoun *mis*. The variables included in this model are dialect, construction type, case and polarity. The response variable in the model is the animacy of the

pronoun. The figure displays all possible splits significant at the level of 0.05 or less. The bar plots at the bottom of the figure show the proportion of inanimate (dark grey) and animate (light grey) observations with the given combination of variable values.

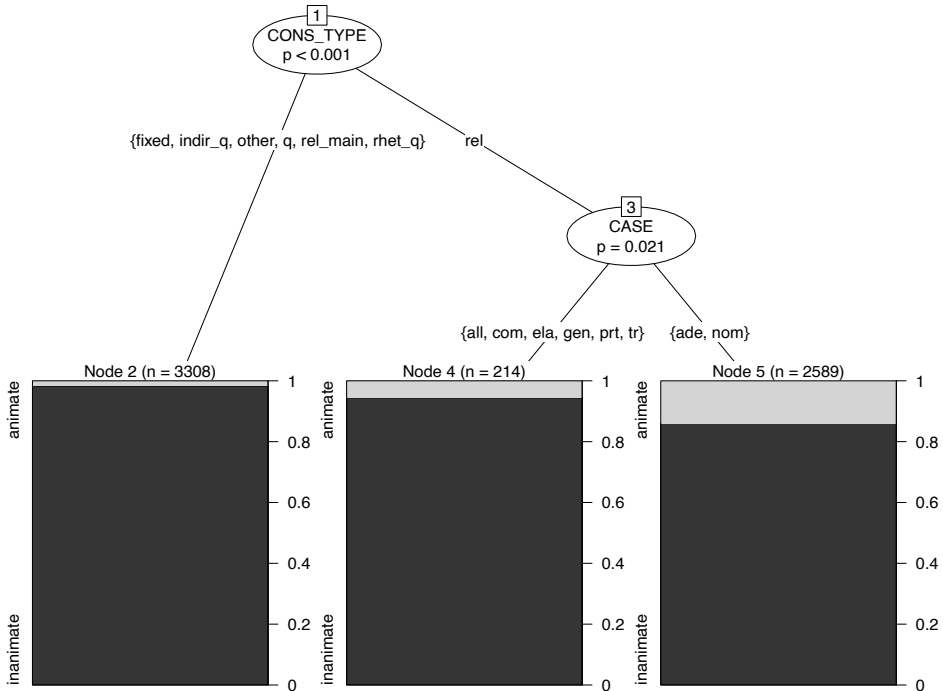


Figure 3. Conditional inference tree for the animacy of the entity to which *mis* refers

According to this model, construction type and case are the variables significantly affecting the speaker's choice for using *mis* to refer to either animate or inanimate entities. The variable of construction type first splits the data into two: using *mis* to refer to animate entities is quite unlikely for all of the construction types (Node 2) except for relative clauses.

The relative clauses are split by the variable of case: it is less likely to refer to animate entities if *mis* is in the allative, comitative, elative, genitive, partitive or translative (Node 4), and more likely if *mis* is either in the nominative or adessive (Node 5). Recall that the previous analysis showed that there were very few instances of the adessive in our dataset, so no fundamental conclusions can be made concerning this case; however, the conditional inference tree supports the previous conclusion that the nominative encourages the speakers' choice to refer to animate entities with *mis*.

Overall, it can be said that *mis* is most likely to refer to an animate being in the Insular, Mid, Mulgi, Northeastern and Western dialects, when *mis* is in the nominative or partitive and in a relative clause (example 9).

- (9) Western (Mar)
see on minu mehe vennatüttar mis sii paergu on
 it be:3SG my husband:GEN niece what here now be:3SG
 ‘This is my husband’s niece, who is here right now.’

5.3. Results from the combined dataset

In this section, I have combined all three datasets for the animacy reference of *kes*, *mis* and *keegi* into one dataset. Table 13 gives the frequencies of the studied pronouns in the CED, separated by dialect. *mis* is best represented in the corpus with 7,558 instances, *kes* follows with 3,709 instances and *keegi* with 1,983 instances. Of all the dialects, these three pronouns appear most frequently in the Mid, Western and Insular dialects. However, as shown in Section 4.1, the total number of tokens from the dialects varies between 48,000 and 251,000, which is, therefore, also reflected in the frequency of the studied pronouns.

Table 13. The frequencies of the pronouns *kes*, *mis* and *keegi* in the CED

| <i>Dialect</i> | <i>kes</i> | <i>mis</i> | <i>keegi</i> | Σ |
|----------------|--------------|--------------|--------------|---------------|
| Mid | 680 | 1,503 | 638 | 2,821 |
| Western | 741 | 1,458 | 511 | 2,710 |
| Insular | 621 | 1,552 | 149 | 2,322 |
| Võru | 389 | 632 | 128 | 1,149 |
| Coastal | 329 | 506 | 114 | 949 |
| Tartu | 191 | 511 | 56 | 758 |
| Mulgi | 187 | 451 | 119 | 757 |
| Eastern | 205 | 323 | 131 | 659 |
| Seto | 195 | 338 | 69 | 602 |
| Northeastern | 171 | 284 | 68 | 523 |
| Σ | 3,709 | 7,558 | 1,983 | 13,250 |

Since the CED consists of spoken texts, a significant number of sentences in the dataset are incomplete or contain repetition of the pronouns. Excluding all the sentences that were impossible or irrelevant to analyse, the final dataset consisted of 3,324 instances of *kes*, 6,111 instances of *mis* and 1,857 instances of *keegi*. Each instance was then annotated with the variables of animacy and polarity (annotation was done according to the same principles as described in Section 5.2). In addition, the variables of case and dialect were acquired from the corpus data.

Table 14 shows the frequency of all the instances of these three pronouns, separated by dialect and their animacy reference.

Table 14. The frequencies of the pronouns *kes*, *mis* and *keegi* in the final dataset, separated by dialect and their animacy reference

| <i>Dialect</i> | <i>kes</i> | | <i>mis</i> | | <i>keegi</i> | |
|----------------|----------------|------------------|----------------|------------------|----------------|------------------|
| | <i>Animate</i> | <i>Inanimate</i> | <i>Animate</i> | <i>Inanimate</i> | <i>Animate</i> | <i>Inanimate</i> |
| Mid | 526 | 87 | 111 | 1,127 | 251 | 344 |
| Western | 588 | 88 | 98 | 1,094 | 176 | 300 |
| Insular | 515 | 36 | 82 | 1,125 | 77 | 61 |
| Võru | 328 | 8 | 27 | 476 | 122 | – |
| Coastal | 256 | 41 | 16 | 391 | 86 | 21 |
| Tartu | 163 | 10 | 25 | 400 | 43 | 9 |
| Mulgi | 145 | 16 | 28 | 343 | 67 | 43 |
| Eastern | 166 | 28 | 11 | 258 | 56 | 72 |
| Seto | 164 | – | 11 | 250 | 63 | – |
| Northeastern | 150 | 9 | 22 | 216 | 46 | 20 |
| Σ | 3,001 | 323 | 431 | 5,680 | 987 | 870 |
| | 3,324 | | 6,111 | | 1,857 | |

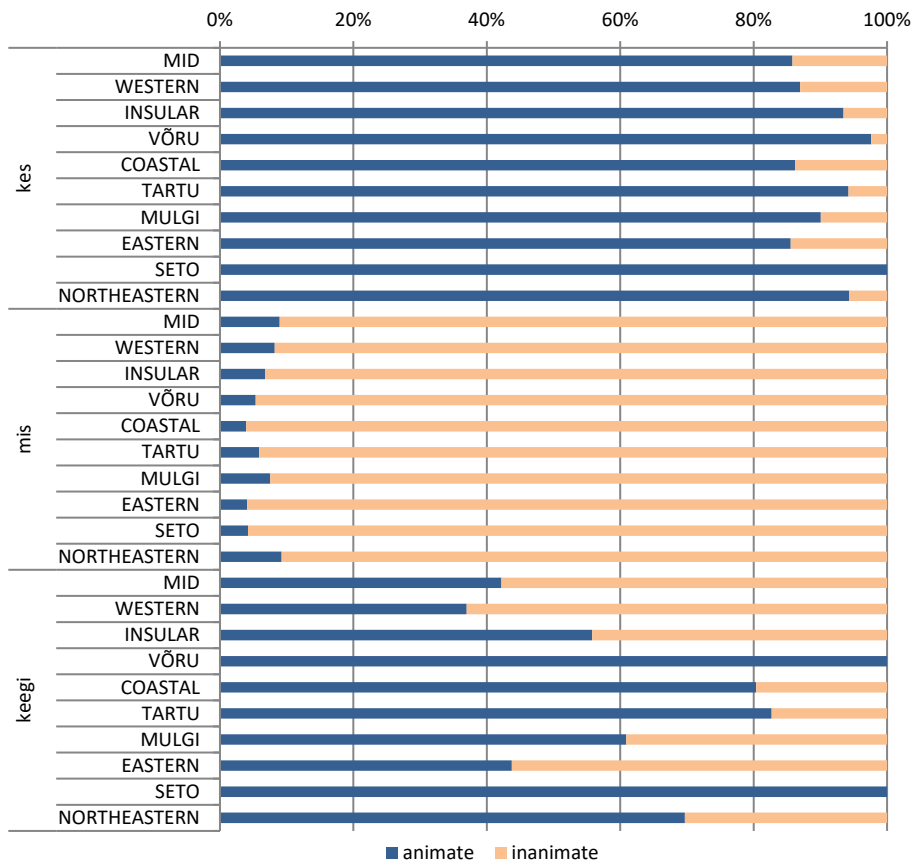


Figure 4. The percentages of the animate and inanimate referents of the pronouns *kes*, *mis* and *keegi* in the final dataset, separated by dialect

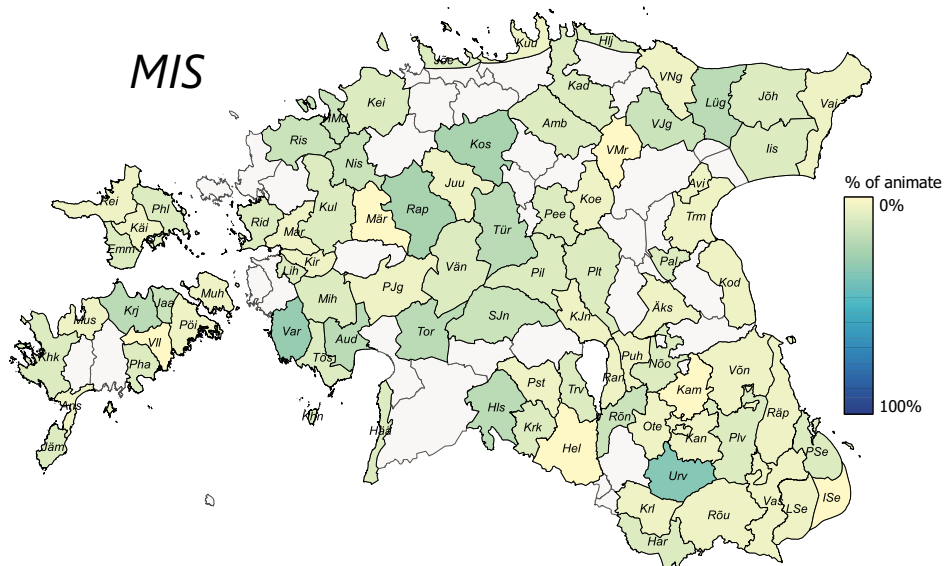


Figure 6. The percentage of the pronoun *mis* used to refer to animate referents in the represented subdialects

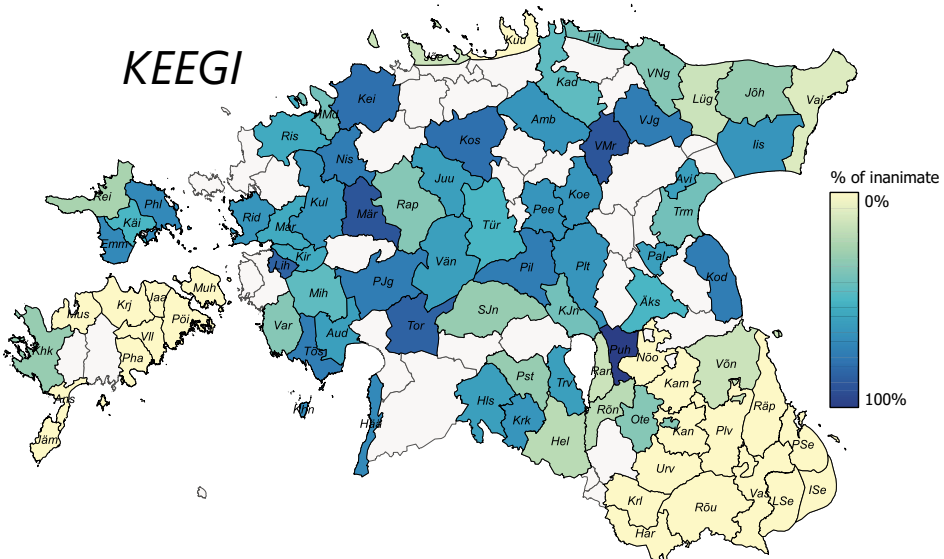


Figure 7. The percentage of the pronoun *keegi* used to refer to inanimate referents in the represented subdialects

Table 15 and Figure 8 show the use of the pronouns according to their case marking. Here, interestingly, the three pronouns all behave in a different manner. The comitative and elative increase the non-selectivity for *kes*, the nominative for *mis* and the partitive for *keegi*. It is also interesting to note that more than half of the comitative and elative forms for *kes* and the partitive forms for *keegi* refer to inanimate entities, but in the case of *mis*, only 7% of the nominative pronouns refer to animate beings. This might indicate that for dialect speakers, the animate use of *mis* is not confined to certain functions or cases and is, instead, used in a varied array of positions.

Table 15. The frequencies of the pronouns *kes*, *mis* and *keegi* in the final dataset, separated by case and animacy reference

| <i>Case</i> | <i>kes</i> | | <i>mis</i> | | <i>keegi</i> | |
|-------------|----------------|------------------|----------------|------------------|----------------|------------------|
| | <i>Animate</i> | <i>Inanimate</i> | <i>Animate</i> | <i>Inanimate</i> | <i>Animate</i> | <i>Inanimate</i> |
| ablative | – | – | – | – | 1 | – |
| adessive | 453 | 15 | 3 | 5 | 156 | 1 |
| allative | 33 | 1 | – | 1 | 35 | – |
| comitative | 44 | 164 | 1 | 92 | 4 | 2 |
| elative | 3 | 29 | – | 32 | – | 6 |
| genitive | 88 | 32 | – | 17 | 39 | 3 |
| inessive | – | – | – | 1 | – | – |
| nominative | 2,322 | 76 | 415 | 5,255 | 587 | 10 |
| partitive | 57 | 6 | 12 | 266 | 165 | 848 |
| translative | 1 | – | – | 11 | – | – |
| Σ | 3,001 | 323 | 431 | 5,680 | 987 | 870 |
| | 3,324 | | 6,111 | | 1,857 | |

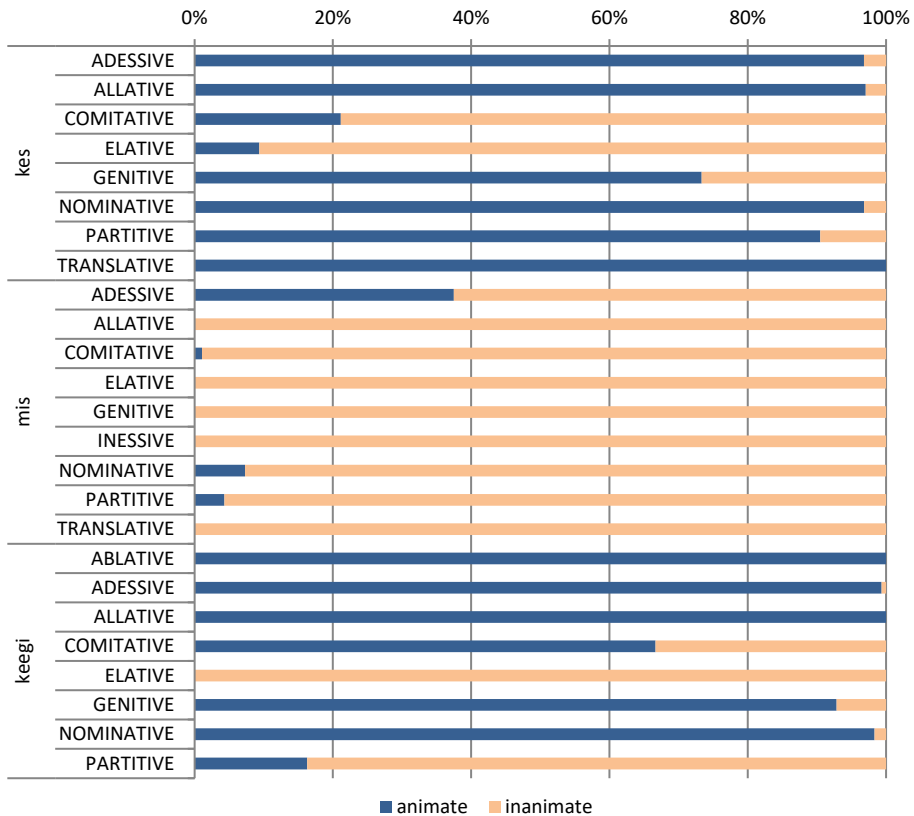


Figure 8. The percentages of the animate and inanimate referents of the pronouns *kes*, *mis* and *keegi* in the final dataset, separated by case

It can be seen from Table 16 and Figure 9 that polarity plays an important role in the use of the pronoun *keegi*, as in 59% of all negative clauses *keegi* refers to an inanimate entity (for affirmative clauses, the percentage is only 17.5%). By contrast, polarity seems to affect the use of the pronoun *kes* in an opposite way: affirmative clauses refer to inanimate entities more frequently (10.3%) than negative clauses (2.6%). It is interesting to see here how the loss of animacy distinction for *keegi* has occurred mainly in the scope of negation, but the same has not happened for *kes*. Moreover, polarity does not seem to bear any effect on the animacy reference of *mis*, as the percentage of animate referents is very similar in both affirmative (7%) and negative (8.5%) clauses.

Table 16. The frequencies of the pronouns *kes*, *mis* and *keegi* in the final dataset, separated by polarity and their animacy reference

| <i>Polarity</i> | <i>kes</i> | | <i>mis</i> | | <i>keegi</i> | |
|-----------------|----------------|------------------|----------------|------------------|----------------|------------------|
| | <i>Animate</i> | <i>Inanimate</i> | <i>Animate</i> | <i>Inanimate</i> | <i>Animate</i> | <i>Inanimate</i> |
| affirmative | 2,775 | 317 | 421 | 5,572 | 448 | 95 |
| negative | 226 | 6 | 10 | 108 | 539 | 775 |
| Σ | 3,001 | 323 | 431 | 5,680 | 987 | 870 |
| | 3,324 | | 6,111 | | 1,857 | |

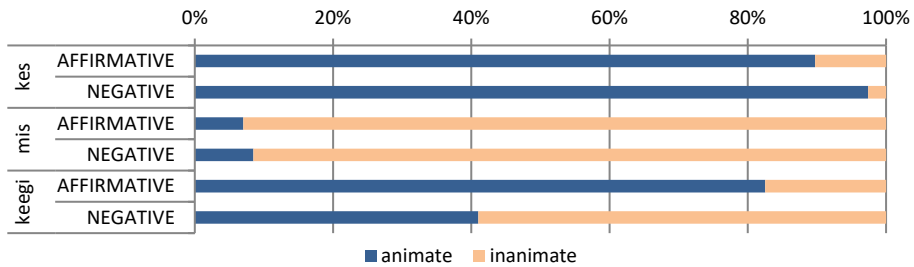


Figure 9. The percentages of the animate and inanimate referents of the pronouns *kes*, *mis* and *keegi* in the final dataset, separated by the polarity of the clause

In order to assess how the chosen variables behave in association with each other, I used multiple correspondence analysis (MCA, see Section 4.2.3), which is displayed in Figure 10. In this model, animacy is not categorised as either animate and inanimate, but rather according to the expected animacy value (*animacy_yes*) and the unexpected animacy value (*animacy_no*) according to Standard Estonian. This means that for *kes* and *keegi*, the expected animacy value is animate and the unexpected animacy value is inanimate. For *mis*, the expected animacy value is inanimate and the unexpected animacy value is animate.

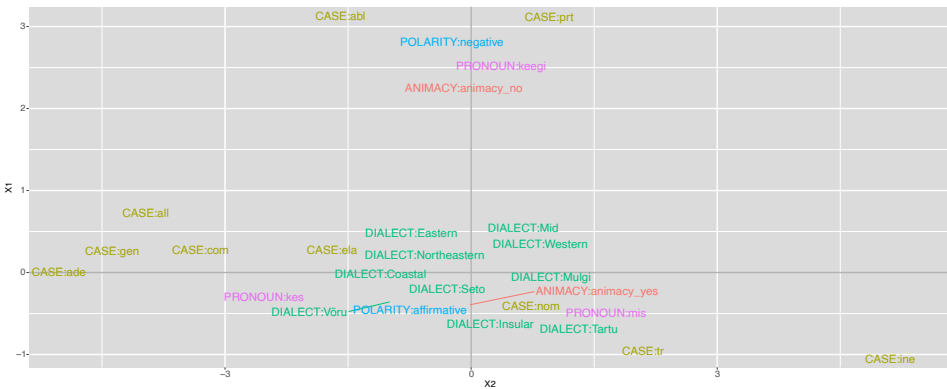


Figure 10. Multiple correspondence analysis for all the variables included in the combined dataset.

The figure shows that the vertical dimension of the graph represents the animacy and polarity scale. The unexpected animacy is strongly associated with negative polarity and the partitive case. The Eastern, Mid, Western and Northeastern dialect are also slightly biased towards the non-selectivity of these three pronouns. It seems that in this context of variables, the partitive case and negative polarity are the most discriminating for this variation, as they are furthest from the origin (the point where the x-axis and y-axis intersect).

Looking at the three pronouns at the focus of this dissertation, we see that the pronouns *kes* and *mis* are more likely to be used to refer to entities of the expected animacy, as the angle connecting them to the expected animacy (animacy_yes) via the origin is quite small. The pronoun *keegi*, by contrast, is very strongly associated with the unexpected animacy. This difference could also be seen in the previous analysis, where the percentage of unexpected animacy was below 10% for both *mis* and *kes*, while *keegi* could be used almost equally to refer to animate and inanimate entities. It is clear that *keegi* has lost much more of its distinction in animacy than either of the interrogative-relative pronouns.

The expected animacy is most associated with affirmative polarity, the nominative case and the southern dialects. However, all these variable levels are rather close to the origin, showing that they might not contribute as strongly to this variation. The huge cluster of variable levels near the origin shows us clearly that the expected animacy (and the variable levels associated with it) is the prototypical and preferred use of the pronouns, as the centre of the plot displays the average use, and, therefore, these variable levels are not as distinct on the figure.

The first dimension of MCA, which represents the animacy and polarity scale, describes 94.2% of the variance in the data. The second dimension describes 4.38% of the variance, so in combination, this MCA analysis describes 98.6% of all the variance in the data. This means that only 1.4% of this particular variation is left unrepresented by this analysis, showing that it is very likely that the most important variables for describing this variation have been included in this analysis.

We can say conclusively that there definitely are similarities in the use of the three pronouns in terms of animacy reference, but they do not behave in identical ways. As for these similarities, we can see that all three pronouns have lost their animacy distinction mostly in the same dialect area: the Eastern, Western and Mid dialects in particular use all the pronouns less selectively than other dialects. The pronouns *mis* and *kes*, which appear in comparable construction types, are also both most likely to reference an entity with an unexpected animacy when they are used in relative clauses.

As for the differences, polarity and case set the pronouns apart. It is clear that the scope of negation is an important factor for the increase of the non-selectivity of *keegi*, as in negated clauses, this pronoun was much more likely to refer to inanimate entities, but the same cannot be said for *kes* and *mis*, for which the polarity of the clause is less or not at all significant. The cases in which the pronouns referred to the entity of an unexpected animacy were different for all three pronouns, indicating that non-selectivity is perhaps tied to certain functions or positions of each pronoun rather than to particular case forms.

One thing that certainly unites all the unexpected uses of the pronouns is grammaticalisation. When either *mis*, *kes* or *keegi* is used in the most grammaticalised position for that particular pronoun, it tends not to distinguish between the animacy of the referent as strictly as it would in a less grammaticalised position. Thus, the speakers most likely use the unexpected pronoun (the one that goes against Standard Estonian norms) when *kes* or *mis* are in the position of a relative pronoun and when *keegi* functions as a negative polarity item. It has been shown for many languages that the further a form moves down the cline of grammaticality, the more it loses its distinction in gender, animacy, number, case, etc. (see Heine & Kuteva 2006: 227). Thus, it can be hypothesised that this increase in non-selectivity for these three pronouns has, at least in part, stemmed from grammaticalisation.

This dissertation has given valuable insight into these previously unexplored variations, but the topic of pronoun animacy reference is by no means exhausted. For example, while it was reasonable in this study to distinguish only two levels of animacy (animate and inanimate), it is possible that dividing animacy into more categories (e.g., whether the referent is a human or an animal, a concrete object or an abstract concept, or whether the referent is expressed in the sentence with a noun, a proper noun or another pronoun) could uncover even clearer patterns of variation.

Another interesting phenomenon apparent in the data is the instances where the speaker uses both *kes* and *mis* in the same or subsequent clause in order to refer to the same referent (examples 10–11). Sometimes the speakers switch to another pronoun in a self-repair clause, but sometimes they use the two pronouns interchangeably without indicating that they have made a mistake. This shows that although there are definitely certain regions, functions and constructions where the variation in animacy reference is more likely to occur, the choice of the pronoun is not always clear even in the speech of a particular speaker. Although the aim of this dissertation was to study this variation from a quantitative perspective, the analysis could potentially benefit from a qualitative approach that takes into account a larger context window than is possible for quantitative analysis.

- (10) Western (Kul)
- | | | | | | | |
|--------------|--------------|---------------|-------------|-------------------|-------------------|-------------|
| <i>arst</i> | <i>jälle</i> | <i>määris</i> | <i>neid</i> | <i>mis</i> | <i>kis</i> | <i>olid</i> |
| doctor | again | rub:PST:3SG | those:PRT | what | who | be:PST:3PL |
| <i>peksa</i> | <i>saand</i> | | | | | |
| beating:PRT | get:PST:PRCL | | | | | |
- ‘The doctor rubbed those again [with medicine] who had been beaten.’

- (11) Coastal (Jõe)
- | | | | | | |
|----------------|-------------|--------------------|-------------|------------|-----------------------|
| <i>teine</i> | <i>mies</i> | <i>miss</i> | <i>siit</i> | <i>oli</i> | <i>kellega</i> |
| other | man | what | from_here | be:PST:3SG | who:COM |
| <i>läksime</i> | | | | | |
| go:PST:1PL | | | | | |
- ‘The other man who was from here [is] who we went with.’

6. CONCLUSION

This dissertation examined the variation in the use of the interrogative-relative pronouns *kes* ‘who’ and *mis* ‘what’ and the indefinite pronoun *keegi* ‘someone, no one’ in Estonian dialects from two syntactic perspectives.

The first variation concerned the animacy reference of the pronouns *kes*, *mis* and *keegi*. In Standard Estonian, *kes* and *mis* and their counterparts *keegi* and *miski* ‘something, nothing’ are distinguished by what they can reference: *kes* and *keegi* are used to refer to animate entities, while *mis* and *miski* are used to refer to inanimate entities. Certain Estonian dialects, however, do not have this restriction and *kes*, *mis* and *keegi* (although not *miski*) can also be used to refer to entities of opposite animacy.

The second variation concerned the object case of the pronoun *mis*. Estonian differentiates between total and partial objects based on semantic and syntactic features: in order to use a total object, the situation has to be perfective, the referent quantitatively bounded and the polarity of the clause affirmative; otherwise, a partial object is required. Total objects occur in nominative or genitive, partial objects in partitive. Nevertheless, there exist a few exceptions to object form variation and the pronoun *mis* is one of them: in a position where a partial object is expected to occur, partitive *mida* can be replaced with nominative *mis*. This variation exists in written, spoken and dialect language.

This dissertation aimed to examine in depth both of these variations in Estonian dialects, in order to determine which geographic and morphosyntactic variables affect speakers’ choice in either the pronoun or object case and discuss the possible motivations behind the emergence of such variations.

The specific research questions this dissertation intended to answer were the following:

- R1. Which Estonian dialects and subdialects typically permit reference to inanimate entities with *kes* and *keegi*, and to animate entities with *mis*?
- R2. Which of the studied variables (clause type, case, polarity, function, position) affect the animacy reference of *kes*, *mis* and *keegi*? Under what conditions does reference to the opposite animacy occur?
- R3. How do the geographic and morphosyntactic variables that affect the animacy reference of *kes*, *mis* and *keegi* compare across these pronouns?
- R4. Which of the studied variables (clause type, polarity, tense, use of impersonal voice, verb type, length of the following word) correlate with the speakers’ choice to use either nominative *mis* or partitive *mida* in partial object position in both Estonian dialects and contemporary speech? Under what conditions does the object case variation happen?
- R5. How do archaic dialects and contemporary speech differ in terms of the variation of *mis* and *mida* in partial object position?
- R6. What are the possible motivations or reasons behind the variation of animacy reference and object case for the pronouns *kes*, *mis* and *keegi*?

This dissertation consists of three published articles and an overview article. P1, P2 and the overview article address the animacy reference of the pronouns *kes*, *mis* and *keegi*, P3 analyses the object case variation of the pronoun *mis*. The data for the studies came from two spoken language corpora: the Corpus of Estonian Dialects and the Phonetic Corpus of Estonian Spontaneous Speech. The former is the basis for all the articles in this dissertation, the latter is used in order to compare the language of archaic dialects to contemporary speech in P3.

R1, R2 and R3 were answered in P1, P2 and the overview article. These studies showed that there are many similarities in animacy reference between the pronouns *kes*, *mis* and *keegi*, but that they also behave in their own unique ways in terms of this variation. It was found that all three pronouns have lost some of their distinction in animacy in similar dialect areas: the Eastern, Western and Mid dialects are most likely to allow reference to inanimate entities with *kes* and *keegi* and to animate entities with *mis*. An increase in this non-selectivity is also seen in the Coastal dialect for *kes* and the Mulgi dialect for *mis*.

A striking difference could be seen between the interrogative-relative and indefinite pronouns: while *mis* and *kes* refer to the opposite animacy in fewer than 10% of the clauses in the data, almost half of the instances of *keegi* refer to inanimate entities. It is clear that *keegi* has lost much more of its animacy distinction than *kes* or *mis*. However, while the percentage of inanimate entities differs greatly, the overall tendencies in the use of *kes* and *keegi* in terms of dialect areas are remarkably similar. For example, the Võru and Seto dialects rarely or never use *kes* or *keegi* to refer to inanimate entities. In addition, the Insular dialect is divided in a comparable manner: *kes* and *keegi* can be used to refer to animates frequently on the island of Hiiumaa, but rarely on the island of Saaremaa. Neither of these tendencies, however, is seen in the use of *mis*.

The variables of polarity and case set the three pronouns apart. The scope of negation was a very important factor for the increase of the non-selectivity of *keegi*, as in negated clauses the pronoun was much more likely to refer to inanimate entities. The same conclusions cannot be drawn for the pronouns *kes* and *mis*, for which the polarity of the clause was much less or not at all significant. While there were certain grammatical cases that speakers preferred for reference to entities of opposite animacy, these cases differed for all three pronouns. This might indicate that the non-selectivity of these pronouns is not as much tied to the case of the pronoun, but rather to the functions the case fills.

Comparing the pronouns *mis* and *kes*, which appear in similar construction types, it was seen that the construction type that mostly encouraged the speakers to refer to the opposite animacy was the relative clause. *keegi* was most likely to refer to inanimate entities when it functioned as a negative polarity item. These results could indicate that the emergence of this non-selectivity is related to grammaticalisation: all three pronouns are most likely to refer to the entity of opposite animacy in their most grammaticalised function – relative clauses for *kes* and *mis*, negative polarity items for *keegi*. As the grammaticalisation of a form is often accompanied by loss of distinction in gender, animacy, number, case, etc. for that

form, grammaticalisation might explain the origins of this variation. This also gives one possible answer to R6 in terms of the variation in animacy reference of *kes*, *mis* and *keegi*.

R4 and R5 were answered in P3. This article showed that the speaker's choice in using nominative *mis* instead of partitive *mida* in the expected partial object position is affected most by verb type, clause type, polarity, length of the following word and dialect. *mida* is more likely to be used in negative relative clauses, with partitive verbs and in the easternmost dialects, while its use was less likely when followed by a monosyllabic pronoun and in (rhetorical) questions. There were significant differences in the percentage of use of the two object cases between archaic dialects and modern speech: in the contemporary data, the use of *mis* and *mida* in the expected partial object position was more or less equal, but in the dialect dataset, partitive *mida* occurred in less than 10% of the sentences.

In order to answer R6 for the object case variation of the pronoun *mis*, it was concluded that there might be multiple motivations behind case variation in partial object position. For example, the lack of general Indo-European or specifically Germanic language contacts in the easternmost dialects and the fact that the north- and southeastern dialects have, over time, been much more conservative in terms of change could explain why the Coastal, Northeastern and Seto dialects have retained the use of partitive *mida*. In addition, very frequent (fixed) constructions consisting of *mis* and a subsequent monosyllabic pronoun showed a higher percentage of use of the nominative than in low-frequency constructions, indicating that high frequency could be the motivation behind the choice to use the shorter nominative in partial object position. Finally, distinct differences in the range of variation in contemporary speech and archaic dialects could be explained by the effects of language standardisation, which has definitely had an influence on contemporary spoken language, but less so on dialects coming primarily from an oral tradition.

Overall, this dissertation has illustrated that all the studied pronouns show variation in their animacy reference, but that this variation occurs in different amounts and that the pronouns can have their own individual variation patterns. A common denominator between the pronouns is that a clear distinction in animacy has become the least relevant for the most grammaticalised functions of the pronouns. The pronouns *kes* and *keegi* share the dialect areas where this variation is most likely to occur, while *mis* does not have as defined of a region of variation. In addition, the articles have endorsed previous claims that in terms of syntactic variation Estonian dialects can not always be divided into northern and southern dialects, but that the largest differences are found instead between the eastern and western dialects.

In conclusion, this dissertation has contributed greatly to the study of the syntactic properties of interrogative-relative and indefinite pronouns as well as to the field of dialect syntax and corpus-based dialectology. It has proven, once again, that variation in language is not random, but governed by a particular set of properties of which speakers might often be unaware, but that are identifiable

by statistical frequency-based analyses. As evidenced by the articles in this dissertation and by previous works on pronouns, it is clear, however, that this field of research contains much more to be studied both in Estonian dialects as well as in contemporary Estonian. It is hoped that this dissertation has helped to bring new perspectives and attention to the complex world of interrogative-relative and indefinite pronoun syntax in Estonian dialects.

SUMMARY IN ESTONIAN

Pronoomenite kasutus ja varieerumine eesti murretes: *kes, mis ja keegi*

Sissejuhatus

Käesolev väitekiri keskendub pronoomenite süntaktilisele varieerumisele eesti murretes. Uurimise all on kolm pronoomenit: küsivad-siduvad asesõnad *kes* ja *mis* ning umbmäärane asesõna *keegi*. Väitekiri uurib nende pronoomenite kasutuse varieerumist kahest vaatenurgast: esiteks on vaatluse all nende pronoomenite võimalus viidata nii elusatele kui ka elututele referentidele (näited 12–14), teisalt pronoomeni *mis* käände varieerumine osaobjekti positsioonis (näited 15–16).

- (12) Western (Rid)
*raha põle kaa **kellega** sa ostma lähäd*
- (13) Mid (Amb)
*sie oli ikke üks nisukkeni inime **miss** ilma maksutta püidis elada*
- (14) Eastern (Pal)
*mes sa vanamis vahid kass kodu **kedagi** amettid eij õle*
- (15) Mid (SJn)
***mida** säääl osta olli*
- (16) Northeastern (Jõh)
*pali õli tämäl seda riijet **miss** nad ise ketrasivad ja kodusivad*

Väitekirjas otsitakse vastust järgnevatele küsimusele:

1. Millised eesti murded ja murrakud lubavad tüüpilisemalt viidata elututele referentidele pronoomenitega *kes* ja *keegi* ning elusatele referentidele pronoomeniga *mis*?
2. Millised uuritud tunnustest (lausetüüp, kääne, kõneliik, funktsioon, positsioon) mõjutavad sellist vastupidise elususega referendile viitamist ja mis tingimustel see varieerumine toimub?
3. Kuidas on seda varieerumist mõjutavad geograafilised ja morfosüntaktilised tunnused võrreldavad uuritud pronoomenite vahel?
4. Millised uuritud tunnustest (lausetüüp, kõneliik, aeg, umbisikuline kõneviis, verbitüüp, järgneva sõna pikkus) korreleeruvad kõnelejate valikuga kasutada kas nominatiivis pronoomenit *mis* või partitiivis *mida* osaobjekti positsioonis ning mis tingimustel see varieerumine toimub?
5. Kuidas erinevad vanapärased murded ja tänapäevane kõne *mis* ~ *mida* varieerumise osas?

6. Mis võivad olla potentsiaalsed põhjused või mõjurid, miks *kes*, *mis* ja *keegi* puhul varieerub viidatava referendi elusus või objekti kääne?

Väitekirjaga koosneb kolmest artiklist ja neile eelnevast ülevaateartiklist. Artiklites P1 ja P2 käsitletakse pronoomenite *kes* ja *keegi* varieerumist referendi elususe suhtes. Artikkel P3 uurib pronoomeni *mis* käände varieerumist eeldatavas osaobjekti positsioonis. Ülevaateartikkel koosneb kuuest peatükist: sissejuhatus, peatükk pronoomenitest ja nende kasutusest, teoreetiliste lähenemisviiside ülevaade, materjali ja meetodi kirjeldus, tulemuste arutelu ning kokkuvõte.

Teoreetiline taust

Elusus. Lingvistiliselt ei ole elusus tavaliselt binaarne tunnus nagu bioloogias, vaid pigem on see kujutletav skaalana, mis algab inimesest ja liigub läbi loomade elutute objektideni (Yamamoto 1999: 1). Elusust keeles kirjeldab elusushierarhia, mille kõige levinum esitus on järgmine (Dixon 1979: 85):

1. ja 2. isiku pronoomen > 3. isiku pronoomen > pärisnimi > inim-NP > elus NP > elutu NP

Mingil kujul elusushierarhia olemasolu peetakse keeltes universaalseks, kuid kõik keeled või grammatilised konstruktsioonid ei pruugi eristada kõikide hierarhia tasemete vahel. Uurimused on näidanud, et kategooriad, mis on hierarhias kõrgemal positsioonil, on tihtipeale grammatiliselt eristatud nendest, mis on hierarhias madalamal, nad on üldises lausestruktuuris tähtsamal positsioonil ja käituvad sagedamini sündmuses agendina. Keeleline elusus võib mõju avaldada laiale valikule nähtustele, nt kääne, ühildumine, arv, sõnajärg, kõneviisi valik, argumentstruktuur jne (Whaley 1996: 172; Kittilä, Västi & Ylikoski 2011: 6; Vihman & Nelson 2019: 261).

Küsiivate-siduvate asesõnade *kes* ja *mis* ning umbmääraste asesõnade *keegi* ja *miski* valik eesti keeles on samuti elusushierarhiast (inimene > kõrgem loom > madalam loom > mitteolend) ning individuaalsushierarhiast (indiviid > distributiivne hulk > kollektiiv). Nendes hierarhiates paremalt vasakule suureneb pronoomenite *kes* ja *keegi* kasutuse tõenäosus ja väheneb *mis* ja *miski* pronoomenite kasutuse tõenäosus. Käesoleva väitekirja üheks lähtekohaks on aga vastupidine tähelepanek, et osades murretes on võimalik pronoomeneid *kes* ja *keegi* kasutada ka elutule referendile ning pronoomenit *mis* elusale referendile viitamiseks.

Grammatiseerumine. Grammatiseerumine on protsess, kus teatud keelevormid muutuvad leksikaalsetest grammatilisteks või grammatilistest veel grammatilistemaks. See muutus toimub samm-haaval pisikeste muutuste kaupa ning võib hõlmata foneeme, sõnu, konstruktsioone ja ka suuremaid diskursuse üksuseid (Lehmann 1985: 303; Heine & Kuteva 2002: 2; Hopper & Traugott 2003: 6–7). Grammatiseerumine on tüüpiliselt ühesuunaline protsess, millel aga ei ole

kindlaksmääratud lõppu. Isegi grammatiseerumise kõige viimases staadiumis võib vorm endiselt olla varieeruvuses ja uuemad ning vanemad vormid võivad kasutuses olla korraga. Tüüpilised grammatiseerumisega kaasnevad protsessid on semantiline pleekimine, konteksti laienemine, degrammatiseerumine ja foneetiline reduktsioon (Lichtenberk 1991: 37; Heine & Kuteva 2002: 2–5; Hopper & Traugott 2003: 16, 67, 127, 138).

Üheks oluliseks teguriks grammatiseerumise protsessis peetakse sagedust: mõned autorid arvavad, et ilma kõrge kasutussageduseta ei saagi grammatiseerumisprotsess üldse alata, samad mõned väidavad, et järsk kasutussageduse kasv märgib tihtipeale grammatiseerumise algust või juba toimunud grammatiseerumist. Mitmed grammatiseerumisele omased muutused (näiteks sõnapiiride hägustumine, erosioon jm) on samuti otseselt seotud vormide kasutamise sagedusega (Heine, Claudi & Hünemeyer 1991: 38–39; Hopper & Traugott 2003: 127; Mair 2004: 125; Heine & Kuteva 2005: 45–46).

Vormide grammatiseerumine on vähemalt osalt olnud ka selles doktoritöös uuritud varieerumise põhjustajaks. Nimelt on küsimarkeritel olnud neli võimalikku grammatiseerumisastet, alustades ainult küsimarkeri funktsioonist kuni selleni, et nende abil on võimalik sisse juhatada põhjaga relatiivlauseid. Erinevates keeltes on näidatud, et sellel skaalal on astmetel alla liikudes vähenenud küsimarkeri täpne kategoriseerimine soo, elususe, arvu, käände jm põhjal ning marker on omandanud järjest üldisema (grammatilise) tähenduse (Heine & Kuteva 2006: 209). See omakorda võib selgitada, miks pronoomeneid *kes* ja *mis* on võimalik kasutada vastupidise elususega referentidele viitamiseks enim just relatiivlausestes.

Dialektoloogia. Dialektoloogia ehk murdeuurimise modernne traditsioon algas Euroopas 19. sajandil. Tavapäraseks murrete kogumise meetodiks kujunesid murdeküsitlused, mille eesmärk oli üldiselt koguda materjali ühtlaselt kõikidelt murdealadelt ja paljudelt erinevatelt keelejuhtidelt. Kuni eelmise sajandi lõpuni keskendus dialektoloogia peamiselt sõnavaralisele ja foneetilisele varieerumisele, kuid morfoloogiliste ja süntaktiliste joonte kogumine ning uurimine oli harv (Kortmann 2010: 843; Szmrecsanyi 2013: 1; Boberg, Nerbonne & Watt 2018: 5–6).

Murdesüntaksi vähesel uurimisel on olnud mitmeid põhjuseid. Näiteks on intervjuudes või küsimustikes üpris raske formuleerida neid küsimusi, millele keelejuht vastaks just täpselt uuritava konstruktsiooniga. Lisaks on takistuseks süntaktilise varieerumise defineerimine, sest see varieerumine tuleb keeles esile pigem sujuvate üleminekutega kasutussagedustena, mitte variandi olemasolu või puudumisena. See omakorda aga tähendab, et süntaktilise varieerumise uurimiseks on vaja kordades rohkem materjali rohkematelt murdealadelt kui mõne muu keelevaldkonna tarbeks (Anderwald & Szmrecsanyi 2009: 1136; Carrilho 2010: 57–58; Kortmann 2010: 843–845, 853).

Alates 1990. aastatest on murdesüntaks aga järjest rohkem keeleteadlaste huviorbiiti tulnud. See on suuresti seotud just salvestusseadmete kasutuselevõtuga ning murdekorpuste koostamise algusega. Korpused on mahukad tekstikogud,

mille abil saab uurida erinevaid lingvistilisi nähtusi ja varieerumisi ning mis võivad koosneda väga erinevast materjalist. Murdekorpused sisaldavad spontaanset (transkribeeritud) suhtlust, mis üldiselt põhineb murdeintervjuudel või keelejuhtide vabal kõnel (Filppula et al. 2005: vii; Kortmann 2010: 844; Szmrecsanyi & Anderwald 2018: 300–301).

Eestis algas süstemaatiline murdematerjali kogumine 1920. aastatel Emakeele Seltsi ja Andrus Saareste eestvedamisel. Pärast patareipõhiste salvestusseadmete kasutuselevõttu 1960. aastatel hakati välitöid murdealadele korraldama igasuviselt. Selle suure töö tulemusena on tänapäeva murdeuurijatel tõeliselt suurepärase kogu murdeülevaateid, murdetekste ja salvestusi, mille põhjal saab teha põhjalikke uurimusi ka tänapäeval, mil autentsete murrete kasutus on enamikes paikades hääbunud (Pajusalu et al. 2020: 35–40; Lindström & Pilvik 2018: 644–648).

Ka Eestis jäi murdeuurimises süntaks esialgu tahaplaanile: üksikud süntaksialased uurimused käsitlesid tihti vaid üht nähtust ühes kitsas murdes või murrakus. Pärast eesti murrete korpuse (EMK) koostamise algust 1998. aastal avardusid võimalused murdesüntaksi uurimiseks ja hakkas kasvama huvi selle murdeuurimise haru vastu. 2013. aastal sai alguse eesti murdesüntaksi projekt (EstDiaSyn¹⁰), mille eesmärgiks sai uurida morfosüntaktilist varieerumist eesti murretes kasutades EMK materjali, rakendades tänapäevaseid statistilisi meetodeid ja pannes erilist rõhku tulemuste visualiseerimisele. Selle projekti raames ja hiljemgi on avaldatud mitmekümneid artikleid ja peetud mitmeid konverentse, mis kõik on oluliselt panustanud meie arusaama eesti murdesüntaksist (Pajusalu et al. 2020: 40; Uiboaed & Lindström 2014: 19; Lindström, Lippus & Tuisk 2019: 345).

Varieerumine. Viimastel aastakümnetel on keele varieerumine ja muutumine saanud üheks kõige produktiivsemaks keeleteaduse uurimisvaldkonnaks. Varieerumise analüüsi aluseks on väide, et eksisteerib kaks või enam viisi öelda üht ja sama asja. Kui varem peeti varieerumist ebaoluliseks ja juhuslikuks, siis üha enam on varieerumise uurimused näidanud, et varieerumine ükskõik millisel keelelisel tasandil ei ole juhuslik, vaid alati mõjutatud mingitest ühiskondlikest või keelelistest faktoritest (Chambers & Trudgill 1998: 49–50; Labov 2004: 7; Bayley 2008: 117, 122; Krug, Schlüter & Rosenbach 2013: 1).

Foneetika või morfoloogia uurimisel on variantide defineerimine üsna ilmselge, kuid sama ei saa öelda süntaksi ja sõnavara puhul. Alati ei ole võimalik kindlalt teada, kas kaks varianti ütlevad täpselt seda sama asja või kas neid siiski eristavad mingid nüansid. Sel juhul kasutatakse tihti neis valdkondades hoopis funktsioonipõhist lähenemist: lingvist määrab kõigepealt uuritava funktsiooni ja leiab siis kõikvõimalikud vormid või konstruktsioonid, mis seda funktsiooni täidavad. Klassikaliseks varieerumise uurimise materjali kogumise meetodiks on intervjuud, mille salvestustelt saab kõikide võimalike variantide sagedusandmeid arvutada (Labov 1972: 72; Feagin 2008: 26; Cameron & Schwenner 2013: 468; Walker 2013: 440–443; Tagliamonte 2012: 9–10).

¹⁰ <https://estdiasyn.ut.ee/project> (Viimati vaadatud 8. veebruar 2023)

Viimased kümnendid on toonud lingvistikasse, sealhulgas ka varieerumise analüüsi kvantitatiivse pöörde ning korpuspõhiste ja tõenäosuslike varieerumiseuurimuste arv on kiirelt kasvanud. Tõenäosuslik lähenemine eeldab, et varieerumine on mõjutatud mitmetest erinevatest piirangutest, mis kõik mõjutavad kõnelejate lingvistilisi valikuid tõenäosuslikult. See omakorda avaldub erinevate variantide statistilises jaotuvuses. Erinevalt traditsioonilisest sotsiolingvistikalisest varieerumise uurimisest, mis keskendub demograafilistele tunnustele nagu vanus, sugu, haridus jne, seostab tõenäosuslik korpuspõhine varieerumine varieerumismustreid ka keelesiseste teguritega nagu sõnavara, vormid, kontekst, abstraktsed piirangud (nt elusus selles doktoritöös) jne (Szmrecsanyi 2017: 688, 693–694; Grafmiller et al. 2018: 1–4; see also Bod, Hay & Jannedy 2003).

Materjal ja meetod

Väitekirja materjal on tulnud kahest korpusest: eesti murrete korpus ja eesti keele spontaanse kõne foneetiline korpus (EKSKFK). Kuna väitekirja puhul on eelkõige tegemist dialektoloogilise uurimisega, siis on EMK olnud sel puhul kõige olulisem lähteandmete allikas: seda on kasutatud kõikides avaldatud artiklites ning lisaks ülevaateartikli peatükis 5.2, kus on uuritud pronoomeni *mis* elususest sõltuvat kasutust, ja peatükis 5.3, kus on kolme uuritud pronoomeni andmed ühendatud ja neid ühiselt analüüsitud. EKSKFK on analüüsi kaasatud artiklis P3 selleks, et võrrelda arhailiste murrete keelt tänapäevase kõnega.

EMK on elektrooniline kollektsioon võimalikult vanapärasest keelematerjalist, mis on kogutud kõigist kümnest eesti murdest. Salvestused on peamiselt tehtud 1960.–1970. aastatel ning keelejuhtidega, kes on üldiselt olnud pigem vanemad naised ja kes ei ole oma elu jooksul kodukülalt kaugemale kolinud. Jututeemadeks on keelejuhtide isiklik elu, kombed, mineviku sündmused jm (Lindström, Lippus & Tuisk 2019: 337–338).

EMKs on eesti murded jaotatud kümneks: põhjatee murdegruppi kuuluvad kesk-, lääne-, ida- ja saarte murre, kirderanniku murdegruppi kuuluvad kirde-murre ja rannamurre ning lõunaeeesti murderühma kuuluvad Tartu, Võru, Mulgi ja Seto murre. Kõik murded on omakorda jaotatud murrakuteks. Korpus koosneb mitmetest kihtidest, alates helisalvestistest kuni keelejuhtide andmebaasini. Selles väitekirjas on kasutatud morfoloogiliselt märgendatud tekste, millest andmestikesse on kaasatud uuritavate pronoomenite kõik esinemisjuhud. Andmestike koostamise ajal sisaldas EMK 1 299 084 morfoloogiliselt märgendatud sõna. Väitekirja andmestikesse on kaasatud 413 teksti 344 keelejuhilt. Pronoomenit *kes* esines lõplikes andmestikes 3324 korda, pronoomenit *mis* 6111 korda ning pronoomenit *keegi* 1857 korda.

Artiklites on kasutatud mitmeid erinevaid statistilised meetodid: klassifitseerimispuud ja -metsad (P1, P2, P3), logistilise regressiooni segamudel (P3) ja mitmene korrespondentsanalüüs (P2). Logistilist regressiooni on varieerumise uurimises rakendatud juba pikemat aega, kuid segamudelite ja klassifitseerimispuude ning -metsade kasutamine on hilisem lähenemine. Uurimused nagu Tagliamonte ja Baayen (2012) ja Baayen jt (2013) on näidanud, et need meetodid

täiendavad üksteist: segamudelites on võimalik arvesse võtta juhuslikke mõjusid, klassifitseerimismetsad näitavad tunnuste olulisust ning klassifitseerimispuud ja korrespondentsanalüüs visualiseerivad tunnuste omavahelisi suhteid.

Tulemused

Doktoritöö esimene artikkel (P1) käsitles pronoomeni *kes* elususest sõltuvat kasutust. Kasutatud murdeandmestikus viitas 9,7% pronoomenitest elututele objektidele, ülejäänud elusatele entiteetidele. Analüüs näitas, et kõige tõenäolisemalt kasutati pronoomenit *kes* elututele objektidele viitamiseks põhjajasti murderühmas (välja arvatud Saaremaal) ja rannamurdes. Lõuna-Eestis esines sellist varieerumist vähem või üldse mitte.

Relatiivlauseid sisaldasid kõige enam elutule viitamist, kuid teistes lausetüüpides oli selline pronoomeni kasutus pigem harv. Elututele objektidele viitamiseks kasutati vaid postnominaalseid relatiivlauseid, samas kui elusale sai viidata nii pre- kui ka postnominaalsega relatiivlausega. See võib olla seotud relatiivmarkeri grammatiseerumisega: kuna see marker funktsioneerib postnominaalsetes relatiivlausetes lihtsalt ühendava lülina ega lisa mingit semantilist väärtust, siis ei ole otseselt oluline, et selle elusus ühtiks ka viidatava referendi elususega. Elututele objektidele viitades oli *kes* peamiselt elatiivis ja komitatiivis, kusjuures nende kahe käändega oli ülekaalukalt viidatud just elutule (90,6% elatiivi puhul ja 78,9% komitatiivi puhul), samas kui kõikide teiste tunnuste kõikide väärtuste puhul oli eelistatud pronoomeni *kes* prototüüpne kasutus.

Doktoritöö teise artikli (P2) fookuses oli umbmäärase pronoomeni *keegi* kasutus. 46,9% kõikidest pronoomeni kasutusjuhtudest andmestikus viitasid elututele referentidele, ülejäänud elusatele referentidele. Kõige sagedamini viidati elutule lääne-, kesk- ja idamurdes, kus üle poolte *keegi* esinemisjuhtudest viitasid elutule. Võru ja Seto murretes ei kasutatud aga pronoomenit elutule viitamiseks kordagi. Kõige tõenäolisem oli elutule referendile viitamine juhul, kui pronoomen oli objekti, partitiivsubjekti või eituslembelise üksuse funktsioonis, partitiivis ja eitava lause lõpus. Kui *keegi* viitas elusale referendile, oli see kõige tõenäolisemalt nominatiivsubjekti funktsioonis ja jaatava lause algul.

Lisaks kirjeldas P2 ka võimalikke pronoomeni *keegi* kasutusfunktsioone murdeandmestikus. Pronoomenit oli võimalik kasutada nii nominatiiv- kui ka partitiivsubjektina, objektina, adverbiaalina, genitiiv- ja järeltäiendina, määratlejana, eituslembelise üksusena ning üldistava alternatiivina. Kõige sagedamini oli *keegi* subjekti funktsioonis, kuid objektid, adverbiaalid ja eituslembelised üksused esinesid samuti tihti.

Ülevaateartikli peatükis 5.2 analüüsiti pronoomeni *mis* elususest sõltuvat kasutust. 7% kõikidest pronoomeni kasutusjuhtudest viitas elusale referendile, ülejäänud viitasid elututele objektidele. Kõige enam kasutati pronoomenit *mis* elusale viitamiseks kirde-, kesk-, lääne ja Mulgi murdes, selline kasutus oli aga harv Seto, ida ja rannamurdes. Sarnaselt pronoomeniga *kes* oli kõige tõenäolisem elusale viidata siis, kui *mis* oli relatiivpronoomen, teistes lausetüüpides oli selline

mitteprototüüpne viitamine harv. Enim viidati elusatele referentidele nominatiivis pronoomeniga. Kõneliik ei osutunud antud analüüsis oluliseks mõjuriks.

Artiklid P1, P2 ja ülevaateartikli peatükid 5.2 ning 5.3 näitasid, et pronoomenitel *kes*, *mis* ja *keegi* on elususest sõltuvas varieerumises nii mõndagi ühist, kuid igaühel on selle varieerumise osas ka eripäraseid käitumismustreid. Uurimuste põhjal võib väita, et kõigil kolmel pronoomenil on elususe eristamine vähenenud sarnastel murdealadel: kõige tõenäolisem on viidata pronoomenitega *kes* ja *keegi* elututele objektidele ja pronoomeniga *mis* elusatele referentidele põhjaeesti murderühmas, eelkõige lääne- ja keskmurdes.

Silmatorfav erinevus küsivate-siduvate ja umbmääraste asesõnade vahel ilmnes nende elutute ja elusate referentide jaotuvuses: kui *kes* ja *mis* viitasid vastupidise elususega referendile vähem kui 10% andmestikus esinevatest lausetest, siis pronoomeniga *keegi* viidati elututele referentidele peaaegu pooltes lausetes. Selge on, et *keegi* puhul on elususe eristus oluliselt rohkem vähenenud võrreldes pronoomenitega *kes* või *mis*. Samas olid üldised levikutendentsid asesõnade *kes* ja *keegi* puhul märkimisväärselt sarnased. Näiteks Võru ja Seto murdealal viitasid need pronoomenid elututele objektidele väga harva või mitte üldse. Samuti oli sarnaselt jaotunud ka saarte murre: pronoomeneid *kes* ja *keegi* kasutati elutule viitamiseks sageli Hiiumaal, kuid harva või mitte kunagi Saaremaal. Selliseid tendentse aga ei olnud näha pronoomeni *mis* puhul.

Kõneliik ja kääne olid kaks tunnust, mis pronoomenite kasutust eristasid. Eituse mõjuala oli väga oluline faktor pronoomeni *keegi* elususe eristuse vähendamisel, sest eitavates lausetes oli palju tõenäolisem, et *keegi* viitab elututele objektidele. Sarnaseid järeldusi ei saanud teha pronoomenite *mis* ja *kes* puhul, mille analüüsid osutus kõneliik vähetähtsaks või ebaoluliseks. Kuigi oli näha, et kõnelejad eelistasid teatud käänetes pronoomenitega viidata vastupidise elususega referendile, siis need konkreetsed käänded erinesid kõigil kolmel pronoomenil. See võib näidata, et elususe eristamine nendel pronoomenitel ei ole niiväga seotud mingi kindla käändega, vaid pigem nende funktsioonidega, mida see kääne või pronoomen konkreetses positsioonis täidab.

Pronoomenite *kes* ja *mis* puhul oli näha, et oluliselt tõenäolisem oli viidata vastupidise elususega referendile siis, kui pronoomen oli relatiivlause. *keegi* viitas kõige sagedamini elutule referendile eituslembelise üksuse funktsioonis. Need tulemused võivad osutada sellele, et selline elususe eristuse hääbumine on seotud grammatiseerumisega, sest kõik kolm pronoomenit viitasid kõige tõenäolisemalt vastupidise elususega referendile just nende kõige enam grammatiseerunud positsioonis: pronoomenite *kes* ja *mis* jaoks on see relatiivlause, *keegi* puhul eituslembeline üksus. Kuna vormide grammatiseerumisega kaasneb tihti ka vähenemine soolises, eluselises, arvulises, käändelises jm eristuses, siis võib grammatiseerumine seletada selle varieerumise algeid.

Doktoritöö kolmas artikkel (P3) uuris pronoomeni *mis* käände varieerumist eeldatavas osaobjekti positsioonis. Artiklis näidati, et kõneleja otsus kasutada selles positsioonis prototüüpse partitiivis *mida* asemel hoopis nominatiivis pronoomenit *mis* on kõige enam seotud verbitüübi, lausetüübi, järgneva sõna pikkuse ja murdetunnusega.

Partitiivis *mida* kasutati suurema tõenäosusega eitavates relatiivlausetes, mis sisaldasid partitiivverbi, samuti idapoolsetes murretes. Nominatiivse *mis* kasutus oli sagedasem, kui seda oli kasutatud küsilauseis või sellele järgnes ühesilbiline (personaal)pronoomen. Artiklis oli eesmärgiks välja selgitada vaid valitud morfo-süntaktiliste tunnuste mõju konkreetsele varieerumisele, ent on tõenäoline, et seda võiks mõjutada ka mõned prosoodiaga seotud tunnused nagu intonatsioon ja lauserõhk.

Tänapäevase kõne ja vananenud murdekeele vahel esinesid pronoomeni *mis* objektikäände varieerumises märkimisväärsed erinevused: tänapäevase kõne andmestikus oli objektikäändena kasutatud nominatiivi ja partitiivi enam-vähem võrdses osakaalus, kuid murdeandmestikus esines partitiivis *mida* vähem kui 10% lausetest.

Artiklis järeldati, et sellisel pronoomeni *mis* käände varieerumisel eeldatavas osaobjekti positsioonis võib olla mitmeid tagamaid. Näiteks võib saksa ja teiste germaani keelekontaktide vähesus idapoolsemates murretes (võrreldes läänepoolsemate murretega) selgitada, miks ranna-, kirde- ja Seto murdes on endiselt säilinud partitiivse *mida* kasutus. Väga sagedastes (kinnis)konstruktsioonides (*mis* + ühesilbiline pronoomen) eelistati kasutada nominatiivi, mis võib osutada sellele, et kõrge kasutussagedus paneb keelekasutajaid valima pigem lühemat nominatiivi. On ka võimalik, et just keele standardiseerimine on põhjustanud suuri osakaalude erinevusi tänapäevase keele ja vananenud murdekeele vahel, sest standardiseerimisel on kindlasti olnud mõju tänapäeva suulisele kõnele, kuid oluliselt vähem murdekeelele.

Kokkuvõte

Siinses doktoritöös uuriti pronoomenite *kes*, *mis* ja *keegi* süntaktilist varieerumist eesti murretes kahest erinevast aspektist: pronoomeniga viidatud referendi elusus ja pronoomeni käände osaobjekti positsioonis. Doktoritöö panustab suuresti nii küsivate-siduvate ja umbmääraste pronoomenite süntaktilise käitumise uurimisse kui ka murdesüntaksi ja korpuspõhise lähenemise valdkonda. Selle doktoritöö (ja ka varasemate pronoomenialaste tööde) põhjal on aga näha, et selles uurimisvaldkonnas on veel palju võimalusi edaspidiseks uurimistööks.

Artiklites tehtud analüüsid ja saadud tulemused näitasid taaskord, et igasugune varieerumine keeles ei ole juhuslik, vaid on mõjutatud erinevatest keelesisestest ja -välistest teguritest. Kõnelejad ei pruugi olla neist teguritest isegi teadlikud, kuid need ilmnevad kvantitatiivsete ja statistiliste sageduspõhiste analüüside käigus. Just (murde)süntaksi uurimisel on selliste meetodite kasutamine eriti oluline, sest vaid nii on võimalik kindlaks teha ka peenemaid kasutusnüansse, mis kvalitatiivses analüüsis ei avalduks.

REFERENCES

- Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory* 21(3). 435–483.
<https://doi.org/10.1023/A:1024109008573>.
- Alberti, Gianmarco. 2015. CAInterprTools: An R package to help interpreting Correspondence Analysis' results. *SoftwareX* 1. 26–31.
<https://doi.org/10.1016/j.softx.2015.07.001>.
- Alvre, Paul. 1977. Pronoomenite morfoloogiat. Pronoomen *kes*. *Keel ja Kirjandus* 1. 18–26.
- Alvre, Paul. 1980. *gi*-liitelisist pronoomeneist. *Keel ja Kirjandus* 9. 539–542.
- Alvre, Paul. 1986. Läänemeresoome pronoomenite tüpoloogiat. *Fenno-Ugristica* 13. 5–20.
- Alvre, Paul. 1987a. Pronoomenite morfoloogiat. Pronoomen *mis*. *Emakeele Seltsi Aastaraamat* 31. 27–36.
- Alvre, Paul. 1987b. *Soome-ugri keelte ajalooline grammatika I. Sõnaliigid. Arvu- ja käändekategooria*. 2. trükk. Tartu: Tartu Riiklik Ülikool.
- Alvre, Paul. 1995. Asesõna *mõni* erijoonetest. *Keel ja Kirjandus* 2. 99–102.
- Anderwald, Lieselotte & Benedikt Szmrecsanyi. 2009. Corpus linguistics and dialectology. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: an international handbook*, vol. 2, 1126–1139. Berlin/New York: Mouton de Gruyter.
- Antso, Siim. 2018. Murdeliigendused XVI–XVIII sajandi eesti kirjalikes allikates. *Emakeele Seltsi Aastaraamat* 63(1). 9–30. <https://doi.org/10.3176/esa63.01>.
- Ariste, Paul. 1956. Eesti asesõnade *me*, *te* ja *meie*, *teie* suhteist. *Keele ja Kirjanduse Instituudi Uurimused* 1. 53–59.
- Ariste, Paul. 1958. Metanalüüsi osa keele arengus. *Keel ja Kirjandus* 11. 681–688.
- Asu, Eva Liina. 2004. The intonation of Saaremaa Estonian: in search of Swedish influence. *Proceedings, Fonetik 2004*. 44–47.
- Asu, Eva Liina. 2005. Intonational Contour Alignment in Saaremaa and Standard Estonian. *Linguistica Uralica* 41(2). 107–112.
- Asu, Eva Liina, Pärtel Lippus, Ellen Niit & Helen Türk. 2012. The acoustic characteristics of monophthongs and diphthongs in the Kihnu variety of Estonian. *Linguistica Uralica* 48(3). 161–170. <https://doi.org/10.3176/lu.2012.3.01>.
- Asu, Eva Liina & Nele Salveste. 2012. The phonetic and phonological analysis of the fall-rise intonation pattern in the Kihnu variety of Estonian. *Linguistica Uralica* 48(3). 171–179.
- Auer, Peter. 2004. Non-standard evidence in syntactic typology – Methodological remarks on the use of dialect data vs spoken language data. In Bernd Kortmann (ed.), *Dialectology meets typology: dialect grammar from a cross-linguistic perspective*, vol. 153, 69–92. Berlin: Mouton De Gruyter.
- Baayen, R. Harald, Douglas J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language* 59(4). 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>.
- Baayen, R. Harald, Anna Endresen, Laura A. Janda, Anastasia Makarova & Tore Nessel. 2013. Making choices in Russian: Pros and cons of statistical methods for rival forms. *Russian linguistics* 37(3). 253–291. <https://doi.org/10.1007/s11185-013-9118-6>.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.
<https://doi.org/10.18637/jss.v067.i01>.
- Bayanati, Shiva & Ida Toivonen. 2019. Humans, animals, things and animacy. *Open Linguistics* 5(1). 156–170. <https://doi.org/10.1515/opli-2019-0010>.

- Bayley, Robert. 2008. The quantitative paradigm. In Jack K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The Handbook of Language Variation and Change* (Blackwell Handbooks in Linguistics), 20–39. Malden: Blackwell.
- Berger, Claudia Bucheli, Elvira Glaser & Guido Seiler. 2012. Is a syntactic dialectology possible? Contributions from Swiss German. In Andrea Ender, Adrian Leemann & Bernhard Wälchli (eds.), *Methods in Contemporary Linguistics* (Trends in Linguistics: Studies and Monographs), 93–120. Berlin/Boston: De Gruyter Mouton.
- Bhat, D. N. S. 2004. *Pronouns* (Oxford Studies in Typology and Linguistic Theory). New York: Oxford University Press.
- Biber, Douglas. 2009. Corpus-based and corpus-driven analyses of language variation and use. In Bernd Heine & Heiko Narrog (eds.), *The Oxford Handbook of Linguistic Analysis*, 159–192. Oxford University Press.
- Björklöf, Sofia. 2012. *Viron rantamurteen länsiryhmän sanaston alkuperä suomalais-lainojen valossa*. Helsinki: University of Helsinki. Master's Thesis.
- Björklöf, Sofia. 2019. Mutual contacts and lexical relations among the Finnic varieties of western Ingria and northeastern Estonia. *Uralica Helsingiensia* 14. 89–153. <https://doi.org/10.33341/uh.85034>.
- Boberg, Charles, John Nerbonne & Dominic Watt. 2018. Introduction. In *The Handbook of Dialectology* (Blackwell Handbooks in Linguistics). Oxford: John Wiley & Sons.
- Bod, Rens, Jennifer Hay & Stefanie Jannedy (eds.). 2003. *Probabilistic linguistics*. Cambridge: MIT Press.
- Breiman, Leo. 2001. Random forests. *Machine learning* 45(1). 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Bouma, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: KNAW.
- Cameron, Richard & Scott Schwenter. 2013. Pragmatics and sociolinguistic variation. In *The Oxford Handbook of Sociolinguistics*, 464–483. Oxford: Oxford University Press.
- Campbell, Lyle. 1991. Some grammaticalization changes in Estonian and their implications. *Approaches to grammaticalization* 1. 285–299.
- Carrilho, Ernestina. 2010. Tools for dialect syntax: the case of CORDIAL-SIN (an annotated corpus of Portuguese dialects). *Tools for linguistic variation*. 57–70. <https://doi.org/10.1387/asju.9891>.
- Chambers, Jack K. & Peter Trudgill. 1998. *Dialectology* (Cambridge Textbooks in Linguistics). 2nd edition. Cambridge: Cambridge University Press.
- Comrie, Bernard. 1989. *Language universals and linguistic typology: Syntax and morphology*. Oxford: Blackwell.
- Croft, William. 1990. *Typology and universals*. Cambridge: Cambridge University Press.
- Crystal, David. 2008. *A dictionary of linguistics and phonetics*. 6th edition. Oxford: Blackwell.
- Dahl, Östen. 2008. Animacy and egophoricity: Grammar, ontology and phylogeny. *Lingua* 118(2). 141–150. <https://doi.org/10.1016/j.lingua.2007.02.008>.
- Dixon, Robert M. W. 1979. Ergativity. *Language* 55(1). 59–138.
- EMS II = Haak, Anu, Evi Juhkam, Varje Lonn, Piret Norvik, Vilja Oja, Valdek Pall & Jüri Viikberg (eds.). 2001. *Eesti murrete sõnaraamat. II köide. 10. vihik: katt-kesä-vuma*. Tallinn: Eesti Teaduste Akadeemia, Eesti Keele Instituut.
- EMS IIIa = Haak, Anu, Evi Juhkam, Varje Lonn, Helmi Neetar, Piret Norvik, Vilja Oja & Jüri Viikberg (eds.). 2001. *Eesti murrete sõnaraamat. III köide. 11. vihik: ket-kiriklik*. Tallinn: Eesti Teaduste Akadeemia, Eesti Keele Instituut.

- EMS IIIb = Haak, Anu, Evi Juhkam, Varje Lonn, Helmi Neetar, Piret Norvik, Vilja Oja & Jüri Viikberg (eds.). 2002. *Eesti murrete sõnaraamat. III köide. 12. vihik: kirikline-kohaline*. Tallinn: Eesti Teaduste Akadeemia, Eesti Keele Instituut.
- EMS VI = Kalvik, Mari-Liis, Mari Kendla & Tiina Laansalu (eds.). 2016. *Eesti murrete sõnaraamat. VI köide. 26. vihik: mehine-muska*. Tallinn: Eesti Teaduste Akadeemia, Eesti Keele Instituut.
- Erelt, Mati. 1996. Relative words in Estonian relative clauses. *Estonian: Typological studies I* (Tartu Ülikooli Eesti Keele Õppetooli Toimetised 4). 9–23.
- Erelt, Mati. 1997. Asesõnadest *ise, enese ~ enda* ja *oma* ning eesti asesõnade süsteemist. *Emakeele Seltsi Aastaraamat* 43. 7–11.
- Erelt, Mati. 2000. On the pro-forms in Estonian. *Estonian: Typological studies IV* (Tartu Ülikooli Eesti Keele Õppetooli Toimetised 14). 7–15.
- Erelt, Mati. 2004. Märkmeid eesti keele komplekslause kohta. *Keel ja Kirjandus* 6. 401–413.
- Erelt, Mati. 2014. *Eesti keele lauseõpetus. Komplekslause*. Tartu: Tartu Ülikooli eesti keele osakond.
- Erelt, Mati. 2017. Komplekslause. Liitlause. In Mati Erelt & Helle Metslang (eds.), *Eesti keele süntaks* (Eesti Keele Varamu III), 647–755. Tartu: Tartu Ülikooli Kirjastus.
- Erelt, Mati, Tiiu Erelt & Kristiina Ross. 2007. *Eesti keele käsiraamat. 3., täiendatud trükk*. Tallinn: Eesti Keele Sihtasutus.
- Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael & Silvi Vare. 1993a. *Eesti keele grammatika I*. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.
- Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael & Silvi Vare. 1993b. *Eesti keele grammatika II*. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.
- Feagin, Crawford. 2008. Entering the community: Fieldwork. In Jack K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The Handbook of Language Variation and Change* (Blackwell Handbooks in Linguistics), 20–39. Malden: Blackwell.
- Fielding, Antony & Harvey Goldstein. 2006. *Cross-classified and multiple membership structures in multilevel models: An introduction and review*. Research Report. Birmingham: University of Birmingham.
- Filppula, Markku, Juhani Klemola, Marjatta Palander & Esa Penttilä. 2005. Introduction. In *Dialects Across Borders: Selected Papers from the 11th International Conference on Methods in Dialectology (Methods XI), Joensuu, August 2002* (Current Issues in Linguistic Theory 273). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Fowler, Roger. 1977. *Linguistics and novel*. London: Routledge.
- Gardelle, Laure & Sandrine Sorlin. 2018. Introduction: Anthropocentrism, egocentrism and the notion of Animacy Hierarchy. *International Journal of Language and Culture* 5(2). 133–162.
- Glynn, Dylan. 2014. Correspondence analysis: exploring data and identifying patterns. In Dylan Glynn & Justyna Robinson (eds.), *Corpus Methods for Semantics. Quantitative studies in polysemy and synonymy*, 343–485. Amsterdam: John Benjamins Publishing Company.
- Goebel, Hans. 1982. *Dialektometrie: Prinzipien und Methoden des Einsatzes der Numerischen Taxonomie im Bereich der Dialektgeographie*. Wien: Österreichische Akademie der Wissenschaften.

- Goebel, Hans. 2006. Recent advances in Salzburg dialectometry. *Literary and linguistic computing* 21(4). 411–435. <https://doi.org/10.1093/lc/fql042>.
- Goebel, Hans. 2011. Dialectometry and quantitative mapping. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language and Space: An International Handbook of Linguistic Variation*, vol. II, 433–457. Berlin: De Gruyter Mouton.
- Grafmiller, Jason, Benedikt Szmrecsanyi, Melanie Röthlisberger & Benedikt Heller. 2018. General introduction: A comparative perspective on probabilistic variation in grammar. *Glossa. A Journal of General Linguistics* 3(1). 94. <https://doi.org/10.5334/gjgl.690>.
- Greenacre, Michael. 2017. *Correspondence analysis in practice*. 2nd edition. New York: Chapman and Hall/CRC.
- Greenacre, Michael J. 1984. *Theory and applications of correspondence analysis*. London: Academic Press.
- Guy, Gregory R. 2005. Letters to Language. *Language* 81(3). 561–563. <https://doi.org/doi:10.1353/lan.2005.0132>.
- Haspelmath, Martin. 2013. Indefinite Pronouns. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://wals.info/chapter/46>.
- Heine, Bernd, Ulrike Claudi & Friederike Hünemeyer. 1991. *Grammaticalization: A conceptual framework*. Chicago/London: University of Chicago Press.
- Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Heine, Bernd & Tania Kuteva. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.
- Heine, Bernd & Tania Kuteva. 2006. *The changing languages of Europe*. Oxford/New York: Oxford University Press.
- Hill, Thomas & Pawel Lewicki. 2006. *Statistics: methods and applications. A comprehensive reference for science, industry, and data mining*. Tulsa: StatSoft, Inc.
- Hint, Helen. 2015. Third person pronoun forms in Estonian in the light of Centering Theory. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 6(2). 105–135. <https://doi.org/10.12697/jeful.2015.6.2.05>.
- Hint, Helen. 2021. *From full phrase to zero: a multifactorial, form-specific and cross-linguistic analysis of Estonian referential system*. Tartu: University of Tartu. PhD Thesis. <http://hdl.handle.net/10062/74058>.
- Hint, Helen, Tiina Nahkola & Renate Pajusalu. 2017. With or without articles? A comparison of article-like determiners in Estonian and Finnish. *Lähivõrdlusi. Lähivertailuja* (27). 65–106. <https://doi.org/10.5128/LV27.02>.
- Hint, Helen, Tiina Nahkola & Renate Pajusalu. 2020. Pronouns as referential devices in Estonian, Finnish, and Russian. *Journal of Pragmatics* 155. 43–63. <https://doi.org/10.1016/j.pragma.2019.10.002>.
- Hint, Helen, Maria Reile & Renate Pajusalu. 2013. Kontekst ja viitamine: argivestlused, legod ja narratiivid. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 4(1). 161–183. <https://doi.org/10.12697/jeful.2013.4.1.10>.
- Hint, Helen, Piia Taremaa, Maria Reile & Renate Pajusalu. 2021. Demonstratiiv-pronoomenid ja -adverbid määratlejatena. Miks me oleme siin ilmas, selles olukorras? *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 12(1). 79–111. <https://doi.org/10.12697/jeful.2021.12.1.03>.

- Hoffmann, Sebastian. 2004. Are low-frequency complex prepositions grammaticalized? On the limits of corpus data – and the importance of intuition. *Corpus Approaches to Grammaticalization in English* 13. 171–210. <https://doi.org/10.1075/scl.13.09hof>.
- Hopper, Paul J. 1991. On some principles of grammaticization. *Approaches to grammaticalization* 1. 17–35.
- Hopper, Paul J. & Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.
- Hothorn, Torsten, Kurt Hornik & Achim Zeileis. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical statistics* 15(3). 651–674. <https://doi.org/10.1198/106186006X133933>.
- Iva, Sulev. 2002. Võro kaudusõ morfofonoloogiast. *Väikeisi kiili kokkoptutmisõq. Väikeste keelte kontaktid* (Võro Instituudi Toimõndusõq 14). 178–184.
- Iva, Sulev. 2003. Võru kõrisulghäälilik. *Lõunaeesti häälilikud II* (Tartu Ülikooli Eesti Keele Õppetooli Toimetised 27). 68–91.
- Iva, Sulev. 2007. *Võru kirjakeele sõnamuutmissüsteem*. Tartu: University of Tartu. PhD Thesis. <http://hdl.handle.net/10062/4860>.
- Iva, Sulev. 2010. Grade Alternation in Võro South Estonian. *Linguistica Uralica* 46(3). 161–174. <https://doi.org/10.3176/lu.2010.3.01>.
- Juhkam, Evi. 1983. Rootsipärane objekt eesti murretes. *Keel ja Kirjandus* 3. 122–125.
- Juhkam, Evi. 2000. Eestirootsi tõlkelaenud eesti murretes. *Inter dialectos nominaque. Pühendusteos Mari Mustale* (Eesti Keele Instituudi Toimetised 7). 53–72.
- Juhkam, Evi. 2012. *Harju-Madise murrak*. Tallinn: Eesti Keele Sihtasutus.
- Juhkam, Evi & Aldi Sepp. 2000. *Läänemurde tekstid* (Eesti Murded VIII). Tallinn: Eesti Keele Instituut.
- Kaiser, Elsi. 2003. *The quest for a referent: A crosslinguistic look at reference resolution*. Philadelphia: University of Pennsylvania. PhD Thesis.
- Kaiser, Elsi. 2010. Effects of contrast on referential form: Investigating the distinction between strong and weak pronouns. *Discourse Processes* 47(6). 480–509. <https://doi.org/10.1080/01638530903347643>.
- Kaiser, Elsi & Katrin Hiietam. 2003. A comparison of the referential properties of third person pronouns in Finnish and Estonian. *Nordlyd* 31(4). 654–667. <https://doi.org/10.7557/12.31>.
- Kaiser, Elsi & Virve-Anneli Vihman. 2010. On the referential properties of Estonian pronouns and demonstratives. In Hans Götzsche (ed.), *Memory, Mind and Language*, 193–205. Newcastle: Cambridge Scholars Publishing.
- Kalmus, Mervi. 2009. *Isiküühildumine eesti murretes*. Tartu: University of Tartu. Master's Thesis.
- Kalvik, Mari-Liis. 2005. Välted muutuvas rannikumurdes. *Keel ja Kirjandus* 3. 209–222.
- Kask, Arnold. 1984. *Eesti murded ja kirjakeel*. Tallinn: Valgus.
- Kask, Arnold & Paula Palmeos. 1985. *Eesti keele grammatika 2: Morfoloogia. Arv- ja asesõnad*. 3. trükk. Tartu: Tartu Riiklik Ülikool.
- Keem, Hella. 1961. Vanapäraseid objektirektsioone Nõo murrakus. *Keel ja Kirjandus* 9. 544–545.
- Keem, Hella. 1970. *Tartu murde tekstid* (Eesti Murded III). Tallinn: Valgus.
- Keem, Hella & Inge Käsi. 2002. *Võru murde tekstid* (Eesti Murded VI). Tallinn: Eesti Keele Instituut.
- Keevallik, Leelo. 1999. The use and abuse of singular and plural address forms in Estonian. *International Journal of the Sociology of Language* 139. 125–144.

- Keevallik, Leelo. 2003. Colloquial Estonian. In Mati Ereht (ed.), *Estonian Language*, 343–374. Tallinn: Estonian Academy Publishers.
- Keevallik, Leelo. 2010. The interactional profile of a placeholder: the Estonian demonstrative *see*. In Nino Amiridze, Boyd Davis & Margaret MacLagan (eds.), *Fillers, pauses and placeholders* (Typological Studies in Language 93), 139–172. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Kendla, Mari. 1997. *Saare- ja Muhumaa kalanimetused*. Tallinn: Eesti Keele Instituut.
- Kendla, Mari. 1999. *Lääne-Eesti kalanimetused*. Tallinn: Eesti Keele Instituut.
- Kendla, Mari. 2005. Eesti liitsõnalised kalanimetused. *Keel ja Kirjandus* 12. 978–983.
- Kettunen, Lauri. 1913a. *Lautgeschichtliche Darstellung über den Vokalismus des Kodaferschen Dialekts* (Suomalais-Ugrilaisen Seuran Toimituksia XXXIV). Helsinki: Suomalais-Ugrilainen Seura.
- Kettunen, Lauri. 1913b. *Lautgeschichtliche Untersuchung über den Kodaferschen Dialekts* (Suomalais-Ugrilaisen Seuran Toimituksia XXXIII). Helsinki: Suomalais-Ugrilainen Seura.
- Kittilä, Seppo, Katja Västi & Jussi Ylikoski. 2011. Introduction to case, animacy and semantic roles. In *Case, Animacy and Semantic Roles* (Typological Studies in Language 99), 1–26. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Kivik, Piibi-Kai. 2010. Personal pronoun variation in language contact: Estonian in the United States. In Muriel Norde, Bob de Jonge & Cornelius Hasselblatt (eds.), *Language contact: New perspectives*, 63–86. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Klavan, Jane, Maarja-Liisa Pilvik & Kristel Uibo. 2015. The use of multivariate statistical classification models for predicting constructional choice in spoken, non-standard varieties of Estonian. *SKY Journal of Linguistics* 28. 187–224.
- Koit, Enn. 1963. Eitus saarte murdes. *Emakeele Seltsi toimetised* 6. *Nonaginta: Johannes Voldemar Veski 90. sünnipäevaks 27. juunil 1963*. 136–147.
- Kont, Karl. 1963. *Käändsõnaline objekt läänemeresoome keeltes*. Tallinn: Eesti NSV Teaduste Akadeemia.
- Koponen, Eino. 1998. *Eteläviron murteen sanaston alkuperä. Itämerensuomalaista etymologiaa* (Suomalais-Ugrilaisen Seuran Toimituksia 230). Helsinki: Suomalais-Ugrilainen Seura.
- Kortmann, Bernd. 2010. Areal variation in syntax. In Peter Auer & Jürgen Erich Schmidt (eds.), *Language and Space: An International Handbook of Linguistic Variation*, 837–864. Berlin/New York: Walter de Gruyter.
- Krikmann, Arvo & Karl Pajusalu. 2000. Kus on keskmurde keskpunkt. *Inter dialectos nominaque. Pühendusteos Mari Mustale* (Eesti Keele Instituudi Toimetised 7). 131–172.
- Kroeger, Paul R. 2005. *Analyzing grammar: An introduction*. New York: Cambridge University Press.
- Krug, Manfred, Julia Schlüter & Anette Rosenbach. 2013. Introduction: Investigating language variation and change. In Manfred Krug & Julia Schlüter (eds.), *Research methods in language variation and change*, 1–14. New York: Cambridge University Press.
- Köpp, Kadri. 2001. *Interrogatiiv-relatiivpronomenid eesti vanemas kirjakeeles*. Tartu: University of Tartu. Bachelor's Thesis.
- Käsi, Inge. 2000. Levinumad adjektiivisufiksids Võru murdes. *Inter dialectos nominaque. Pühendusteos Mari Mustale* (Eesti Keele Instituudi Toimetised 7). 173–178.

- Laanesoo, Kirsi. 2013. Direktiivsed *mis*-ja *mida*-küsilauseid suulises suhtluses. *Emakeele Seltsi Aastaraamat* 59. 103–126. <https://doi.org/10.3176/esa59.05>.
- Labov, William. 1963. The social motivation of a sound change. *Word* 19(3). 273–309.
- Labov, William. 1966. *The social stratification of English in New York City*. Washington: Center for Applied Linguistics.
- Labov, William. 1972. *Sociolinguistic patterns* (Conduct and Communication 4). Philadelphia: University of Pennsylvania Press.
- Labov, William. 2004. Quantitative analysis of linguistic variation. In Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier & Peter Trudgill (eds.), *Sociolinguistics: An International Handbook of the Science of Language and Society*, vol. I, 6–21. Berlin: Mouton de Gruyter.
- Lamers, Monique & Peter de Swart (eds.). 2012. *Case, word order and prominence: Interacting cues in language production and comprehension* (Studies in Theoretical Psycholinguistics 40). Dordrecht: Springer.
- Langacker, Ronald W. 1991. *Foundations of cognitive grammar. Volume II: Descriptive application*. Stanford: Stanford University Press.
- Lehmann, Christian. 1985. Grammaticalization: Synchronic variation and diachronic change. *Lingua e Stile* 20. 303–318.
- Lichtenberk, Frantisek. 1991. On the gradualness of grammaticalization. *Approaches to grammaticalization* 1. 37–80.
- Lindström, Liina. 1997. Eitus Võru murde suulises kõnes. In *Läänemeresoome lõunapiir. Võro instituudi toimõtisõq*, 143–154. Võro: Võro Instituut.
- Lindström, Liina. 2000. Narratiiv ja selle sõnajärg. *Keel ja Kirjandus* 3. 190–200.
- Lindström, Liina. 2001a. Verb-initial clauses in narrative. *Estonian: Typological studies V* (Tartu Ülikooli Eesti Keele Õppetooli Toimetised 18). 138–168.
- Lindström, Liina. 2001b. Eesti murrete korpuse iseloomustus argivestlustega võrrelduna. *Keele kannul. Pühendusteos Mati Erealti 60. sünnipäevaks 12. märtsil 2001* (Tartu Ülikooli Eesti Keele Õppetooli Toimetised 17). 212–221.
- Lindström, Liina. 2006. Infostruktuuri osast eesti keele sõnajärje muutumisel. *Keel ja Kirjandus* 11. 875–888.
- Lindström, Liina. 2017. Partitive subjects in Estonian dialects. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 8(2). 191–231. <https://doi.org/10.12697/jeful.2017.8.2.07>.
- Lindström, Liina, Liisi Bakhoff, Mari-Liis Kalvik, Anneliis Klaus, Rutt Läänemets, Mari Mets, Ellen Niit, Karl Pajusalu, Pire Teras, Kristel Uihoaed, Ann Veismann & Eva Velsker. 2006. Sõnaliigituse küsimusi eesti murrete korpuse põhjal. *Keele ehe* (Tartu Ülikooli Eesti Keele Õppetooli Toimetised 30). 154–167.
- Lindström, Liina, Mervi Kalmus, Anneliis Klaus, Liisi Bakhoff & Karl Pajusalu. 2009. Ainsuse 1. isikule viitamine eesti murretes. *Emakeele Seltsi aastaraamat* 54. 159–185.
- Lindström, Liina, Pärtel Lippus & Tuuli Tuisk. 2019. The online database of the University of Tartu Archives of Estonian Dialects and Kindred Languages and the Corpus of Estonian Dialects. *Uralica Helsingiensia* 14. <https://doi.org/10.33341/uh.85040>.
- Lindström, Liina, Varje Lonn, Mari Mets, Karl Pajusalu, Pire Teras, Ann Veismann, Eva Velsker & Jüri Viikberg. 2001. Eesti murrete korpus ja kolme murde sagedasema sõnavara võrdlus. *Keele kannul. Pühendusteos Mati Erealti 60. sünnipäevaks 12. märtsil 2001* (Tartu Ülikooli Eesti Keele Õppetooli Toimetised 17). 186–211.
- Lindström, Liina & Karl Pajusalu. 2003. Corpus of Estonian dialects and the Estonian vowel system. *Linguistica Uralica* 4. 241–257.

- Lindström, Liina & Maarja-Liisa Pilvik. 2018. Korpuspõhine kvantitatiivne dialektoloogia. *Keel ja Kirjandus* 8–9. 643–662. <https://doi.org/10.54013/kk730a3>.
- Lindström, Liina, Maarja-Liisa Pilvik & Helen Plado. 2018. Nimetamiskonstruktsioonid eesti murretes: murdeerinevused või suuline süntaks? *Mäetagused* 70. 91–126. https://doi.org/10.7592/MT2018.70.lindstrom_pilvik_plado.
- Lindström, Liina, Maarja-Liisa Pilvik & Helen Plado. 2021. Variation in negation in Seto. *Studies in Language*. John Benjamins 45(3). 557–597. <https://doi.org/dx.doi.org/10.1075/sl.19063.lin>.
- Lindström, Liina, Maarja-Liisa Pilvik, Mirjam Ruutma & Kristel Uiboaed. 2015. Mineviku liitaegade kasutusest eesti murretes keelekontaktide valguses. *Võro Instituudi toimõndusõq* 29. 39–70.
- Lindström, Liina, Maarja-Liisa Pilvik, Mirjam Ruutma & Kristel Uiboaed. 2019. On the use of perfect and pluperfect in Estonian dialects. *Uralica Helsingiensia* 14. 155–193. <https://doi.org/10.33341/uh.85035>.
- Lindström, Liina & Kristel Uiboaed. 2017. Syntactic variation in ‘need’-constructions in Estonian dialects. *Nordic Journal of Linguistics* 40(3). 313–349. <https://doi.org/10.1017/S0332586517000191>.
- Lindström, Liina, Kristel Uiboaed & Virve-Anneli Vihman. 2014. Varieerumine *tarvis-/vaja*-konstruktsioonides keelekontaktide valguses. *Keel ja Kirjandus* 8–9. 609–630. <https://doi.org/10.54013/kk682a4>.
- Lindström, Liina, Eva Velsker, Ellen Niit, Karl Pajusalu, Marja Kallasmaa & Vilja Oja. 2009. *Mees* ‘man’, *aeg* ‘time’ and other frequent words in the Corpus of Estonian Dialects. In Marja Kallasmaa & Vilja Oja (eds.), *Kodukeel ja keele kodu/Home language and the home of a language* (Eesti Keele Instituudi Toimetised 13), 91–129. Tallinn: Eesti Keele Sihtasutus.
- Lindström, Liina & Virve-Anneli Vihman. 2010. *Ise ise. Eesti ja soome-ugri keeleteaduse ajakiri*. *Journal of Estonian and Finno-Ugric Linguistics* 1(2). 219–241.
- Lippus, Pärtel. 2003. Võru afrikaadid. *Lõunaeesti häälikud II* (Tartu Ülikooli Eesti Keele Õpetooli Toimetised 27). 115–130.
- Lippus, Pärtel, Kätlin Aare, Anton Malmi, Tuuli Tuisk & Pire Teras. 2021. Phonetic Corpus of Estonian Spontaneous Speech v1.2. Institute of Estonian and General Linguistics, University of Tartu. <https://doi.org/10.23673/RE-293>.
- Lonn, Varje & Ellen Niit. 2002. *Saarte murde tekstid* (Eesti Murded VII). Tallinn: Eesti Keele Instituut.
- Mair, Christian. 2004. Corpus linguistics and grammaticalisation theory. *Corpus Approaches to Grammaticalization in English* 13. 121–150.
- McEnery, Tony & Andrew Hardie. 2011. *Corpus linguistics: Method, theory and practice*. New York: Cambridge University Press.
- Mets, Mari. 2007. South Estonian written standard and actual spoken language: variation of the past participle markers. *Linguistica Uralica* 43(3). 161–172.
- Mets, Mari. 2010. *Suhtlusvõrgustikud reaalsajas: võru kõnekeele varieerumine kahes Võrumaa külas*. Tartu: University of Tartu. PhD Thesis. <http://hdl.handle.net/10062/14781>.
- Mets, Mari. 2011. Spoken Võro in real time. Variation of the inessiv ending. *Linguistica Uralica* 47(4). 257–272. <https://doi.org/10.3176/lu.2011.4.02>.
- Mets, Mari. 2013. Võru kõnekeel reaalsajas. *tud*-partitsiibi tunnuse varieerumine. *Keel ja Kirjandus* 12. 900–909. <https://doi.org/10.54013/kk673a2>.
- Metslang, Helle. 2003. *-ki/gi, ka* ja nende soome kaimud. *Lähivertailuja* 12. 57–81.

- Metslang, Helle. 2017. Lauseliikmed. Sihitis. In Mati Ereht & Helle Metslang (eds.), *Eesti keele süntaks* (Eesti Keele Varamu III), 258–277. Tartu: Tartu Ülikooli Kirjastus.
- Metslang, Helle & Liina Lindström. 2017. Essive in Estonian. *Uralic essive and the expression of impermanent state* 119. 57–90. <https://doi.org/10.1075/tsl.119.03met>.
- Metslang, Helle & Karl Pajusalu. 2002a. Evidentiality in South Estonian. *Linguistica Uralica* 38(2). 98–109.
- Metslang, Helle & Karl Pajusalu. 2002b. Evidentsiaalsusest eesti keeles ja Tartu murde *na*-tunnuselisest kaudest kõneviisist. *Väikeisi kiili kokkupuutmisõq. Väikeste keelte kontaktid*. (Võro Instituudi Toimõndusõq 14). 161–177.
- Metsmägi, Iris, Meeli Sedrik & Vilja Oja. 2014. Different loan sources as a reason for variability of some Estonian dialect words. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 5(2). 51–70. <https://doi.org/10.12697/jeful.2014.5.2.03>.
- Metsmägi, Iris, Meeli Sedrik & Sven-Erik Soosaar. 2012. *Eesti etimoloogiasõnaraamat*. Tallinn: Eesti Keele Sihtasutus.
- Milroy, James & Lesley Milroy. 1997. Varieties and variation. In Florian Coulmas (ed.), *The Handbook of Sociolinguistics* (Blackwell Handbooks in Linguistics 4). Oxford/Cambridge: Blackwell.
- Murumets, Sirje. 1982. Eesti keeleala murdelisest liigendusest “Väikese murdesõnastiku” põhjal, I. *Keel ja Kirjandus* 1. 11–17.
- Murumets, Sirje. 1983. Eesti keeleala murdelisest liigendusest “Väikese murdesõnastiku” põhjal, II. *Keel ja Kirjandus* 11. 615–623.
- Murumets, Sirje. 1984. Regional Vocabulary and Lexical Regions in Estonia. In Hans Goebel (ed.), *Dialectology* (Quantitative Linguistics 21), 224–253. Bochum: Brockmeyer.
- Must, Mari. 1987. *Kirderannikumurre: häälikuline ja grammatiline ülevaade*. Tallinn: Valgus.
- Must, Mari. 1995. *Kirderannikumurde tekstid* (Eesti Murded V). Tallinn: Eesti Keele Instituut.
- Must, Mari. 2000. *Vene laensõnad eesti murretes*. Tallinn: Eesti Keele Sihtasutus.
- Must, Mari & Aili Univere. 2002. *Põhjaeesti keskmurre. Häälikulisi ja morfoloogilisi peajooni* (Eesti Keele Instituudi Toimetised 10). Tallinn: Eesti Keele Instituut.
- Mäger, Mart. 1967. *Eesti linnunimetused*. Tallinn: Keele ja Kirjanduse Instituut.
- Neetar, Helmi. 1963. Ühest aluse ja öeldise ühildumise juhust eesti murretes. *Emakeele Seltsi toimetised 6. Nonaginta: Johannes Voldemar Veski 90. sünnipäevaks 27. juunil 1963*. 217–223.
- Neetar, Helmi. 1964. Aluse ja öeldise ühildumist mõjutavatest teguritest eesti murretes. *Emakeele Seltsi Aastaraamat* 10. 151–166.
- Neetar, Helmi. 1965a. Aluse (kollektiivsubstantiivi) ja öeldise ühildumisest eesti murretes. *Emakeele Seltsi Aastaraamat* 11. 185–193.
- Neetar, Helmi. 1965b. Aluse ja öeldise ühildumise seaduspärasustest eesti murretes. *Keel ja Kirjandus* 1. 25–29.
- Neetar, Helmi. 1973. *i*-mitmusest Hiiumaal. *Emakeele Seltsi Aastaraamat* 19–20. 205–217.
- Nenadic, Oleg & Michael Greenacre. 2007. Correspondence analysis in R, with two- and three-dimensional graphics: the *ca* package. *Journal of statistical software* 20(3). 1–13. <https://doi.org/10.18637/jss.v020.i03>.
- Nerbonne, John & Peter Kleiweg. 2007. Toward a dialectological yardstick. *Journal of Quantitative Linguistics* 14(2–3). 148–166. <https://doi.org/10.1080/09296170701379260>.

- Nevis, Joel A. 1984. A non-endoclititic in Estonian. *Lingua* 64(2–3). 209–224.
- Niit, Ellen. 2005. Vowels in initial syllables in Saaremaa. *Linguistica Uralica* 41(2). 113–122.
- Niit, Ellen. 2007. Vowels in word-initial syllables in Muhu Estonian. *Linguistica Uralica* 43(4). 241–249.
- Niklus, Lea. 1957. *Otepää murrak*. Tartu: University of Tartu. Diploma Thesis.
- Norvik, Piret. 1985. *On the lexical relations between the North-Estonian dialects*. Tallinn: Academy of Sciences of the Estonian SSR.
- Norvik, Piret. 2000. Kirderannikumurde sõnavaralisi suhteid. *Inter dialectos nominaque. Pühendusteos Mari Mustale* (Eesti Keele Instituudi Toimetised 7). 184–201.
- Nurkse, Rein. 1937. *Adjektiiv-atribuudi kongruentsist eesti keeles*. Tartu: Akadeemilise Emakeele Seltsi Kirjastus.
- Ogren, David. 2015. Differential object marking in Estonian: Prototypes, variation, and construction-specificity. *SKY Journal of Linguistics* 28. 277–312.
- Oja, Vilja. 1996. Compound adjectives used in the Estonian dialects to denote light and pale shades of colour. *Linguistica Uralica* 32(2). 92–100.
- Oja, Vilja. 2001. *Linguistic studies of Estonian colour terminology*. Tartu: University of Tartu. PhD Thesis.
- Org, Laivi. 2003. Vastseliina palatalisatsioonist. *Lõunaestli häälikud II* (Tartu Ülikooli Eesti Keele Õpetooli Toimetised 27). 131–148.
- Org, Laivi. 2005. *Võru palatalisatsiooni akustikast*. Tartu: University of Tartu. Master's Thesis. <http://hdl.handle.net/10062/819>.
- Org, Laivi. 2016. Võru keele palatalisatsiooni akustikast. *Õdagumeresuumlaisi nimeq. Läänemeresoomlaste nimed* (Võru Instituudi Toimõtisõq 30). 239–273.
- Pajusalu, Karl. 1989. Suhtlustaotlused inimkeelt muutmas. Tähelepanekuid eesti murrete verbimorfoloogiast. *Keel ja Kirjandus* 3. 142–149.
- Pajusalu, Karl. 1996. *Multiple linguistic contacts in South Estonian: Variation of verb inflection in Karksi*. Turku: University of Turku.
- Pajusalu, Karl. 1998. Vowel reduction in south Estonian. *Linguistica Uralica* 34(3). 234–240.
- Pajusalu, Karl. 2005. Keelesaarte *se-minevik*. *Tartu Ülikooli Lõuna-Eesti keele-ja kultuuriuuringute keskuse aastaraamat IV*. 145–150.
- Pajusalu, Karl, Tiit Hennoste, Ellen Niit, Peeter Päll & Jüri Viikberg. 2020. *Eesti murded ja kohanimed*. 3., kohendatud ja täiendatud trükk. Tallinn: Eesti Keele Sihtasutus.
- Pajusalu, Karl, Arvo Krikmann & Eberhard Winkler. 2009. Lexical relations between Salaca Livonian and Estonian dialects. *Linguistica Uralica* 45(4). 283–298. <https://doi.org/10.3176/lu.2009.4.04>.
- Pajusalu, Karl & Pire Teras. 2012. Ühest edelaläänemeresoome fonoloogilisest ühisuendusest: prevokalisatsioon eesti ja liivi murretes. *Pühendusteos emeritprofessor Mati Hindi 75. sünnipäevaks* (Tallinna Ülikooli Eesti Keele Ja Kultuuri Instituudi Toimetised 14). 156–176.
- Pajusalu, Karl, Eva Velsker & Ervin Org. 1999. On recent changes in South Estonian: dynamics in the formation of the inessive. *International Journal of the Sociology of Language* 139. 87–103.
- Pajusalu, Renate. 1996. Pronoun systems of common Estonian and Estonian dialects in a contrastive perspective. *Estonian: Typological Studies I* (Tartu Ülikooli Eesti Keele Õpetooli Toimetised 4). 145–164.
- Pajusalu, Renate. 1997a. Eesti pronomeneid I/1. Ühiskeele *see, too* ja *tema/ta*. *Keel ja Kirjandus* 1. 24–30.

- Pajusalu, Renate. 1997b. Eesti pronomeneid I/2. Ühiskeele *see, too* ja *tema/ta*. *Keel ja Kirjandus* 2. 106–115.
- Pajusalu, Renate. 1997c. Is there an article in (spoken) Estonian? *Estonian: Typological studies* II (Tartu Ülikooli Eesti Keele Õpetooli Toimetised 8). 146–177.
- Pajusalu, Renate. 1998. Eesti pronomeneid II. Võru *sjoo, taa, tuu* ja *timä*. *Keel ja Kirjandus* 3. 159–172.
- Pajusalu, Renate. 2000. Indefinite determiners *mingi* and *üks* in Estonian. *Estonian: Typological studies* IV (Tartu Ülikooli Eesti Keele Õpetooli Toimetised 14). 87–117.
- Pajusalu, Renate. 2001. Definite and indefinite determiners in Estonian. *Pragmatics in 2000: Selected papers from the 7th International Pragmatics Conference Antwerp: International Pragmatics Association*. 458–469.
- Pajusalu, Renate. 2004. Viron *üks* ja *kõik*. *Virittäjä* 108(1). 2–23.
- Pajusalu, Renate. 2005. Anaphoric pronouns in spoken Estonian: crossing the paradigms. In Ritva Laury (ed.), *Minimal reference. The use of pronouns in Finnish and Estonian discourse*, 107–134. Helsinki: Finnish Literature Society.
- Pajusalu, Renate. 2006a. Death of a demonstrative: person and time. The case of Estonian *too*. *Linguistica Uralica* 42(4). 241–253. <https://doi.org/10.3176/lu.2006.4.01>.
- Pajusalu, Renate. 2006b. Multiple motivations for meaning of an elative *wh*-construction in Estonian. *Trames* 10(4). 324–340. <https://doi.org/10.3176/tr.2006.4.02>.
- Pajusalu, Renate. 2008. Pragmatics of quantifiers: The case of Estonian *kõik* ‘all’. *Journal of Pragmatics* 40(11). 1950–1965. <https://doi.org/10.1016/j.pragma.2008.04.006>.
- Pajusalu, Renate. 2009. Pronouns and reference in Estonian. *Language Typology and Universals* 62(1–2). 122–139. <https://doi.org/10.1524/stuf.2009.0008>.
- Pajusalu, Renate. 2010. *Teie* või *sina*? *Oma keel* 1. 5–10.
- Pajusalu, Renate. 2014. Palved eesti, soome ja vene keeles: grammatika pragmaatika teenistuses. *Eesti Rakenduslingvistika Ühingu aastaraamat* 10. 241–257. <https://doi.org/10.5128/ERYa10.15>.
- Pajusalu, Renate. 2015. Võro demonstratives: changing or disappearing? *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 6(2). 167–190. <https://doi.org/10.12697/jeful.2015.6.2.07>.
- Pajusalu, Renate. 2017. Lause kommunikatiivne struktuur ja süntaktilised protsessid. Viiteseosed. In Mati Ereht & Helle Metslang (eds.), *Eesti keele süntaks* (Eesti Keele Varamu III), 566–589. Tartu: Tartu Ülikooli Kirjastus.
- Pajusalu, Renate. 2019. Viisakusteooria ja 100 aastat eesti viisakust: väike sissevaade Lutsu ja Tammsaare teoste põhjal. In Riho Altnurme (ed.), *Humanitaarteadused ja kunstid 100-aastases rahvusülikoolis*, 132–145.
- Pajusalu, Renate, Birute Klaas-Lang, Karl Pajusalu, Andra Kalnača & Ilze Lokmane. 2017. Requests, questions and space: evidence from Estonian, Latvian, Lithuanian, Finnish and Russian. *Valoda: nožme un forma* 8. 212–231. <https://doi.org/10.22364/VNF.8.16>.
- Pajusalu, Renate, Maria Reile, Helen Hint, Tiina Nahkola & Piia Taremaa. 2018. Relative clauses in spatial and narrative contexts in Estonian, Finnish, and Russian. *SKY Journal of Linguistics* 31. 107–142.
- Pajusalu, Renate, Virve-Anneli Vihman, Birute Klaas & Karl Pajusalu. 2010a. Forms of address across languages: Formal and informal second person pronoun usage among Estonia’s linguistic communities. *Intercultural Pragmatics* 7(1). 75–101. <https://doi.org/10.1515/IPRG.2010.004>.

- Pajusalu, Renate, Virve-Anneli Vihman, Birute Klaas & Karl Pajusalu. 2010b. Eestlaste ja venelaste suhtluskäitumine: *sina, teie* ja keegi veel. *Eesti Rakenduslingvistika Ühingu aastaraamat* 6. 207–224. <https://doi.org/10.5128/ERYa6.13>.
- Paldre, Leho. 1998. *Eitustundlikud üksused eesti keeles*. Tartu: University of Tartu. Master's Thesis.
- Pall, Valdek. 1994. Sissejuhatus. Eesti murretest. In *Eesti murrete sõnaraamat. I köide. 1. vihik*, 5–12. Tallinn: Eesti Keele Instituut.
- Pant, Annika. 2018. Üle saja aasta õigekeelsusprobleeme: *gi-/ki*-liite paiknemine aseenade käändevormides. *Oma Keel* 2. 29–36.
- Pant, Annika. 2020. *Pronoomenite keegi, miski, kumbki, ükski käändevormide kasutus tänapäeva eesti keeles*. Tartu: University of Tartu. Master's Thesis. <http://hdl.handle.net/10062/68143>.
- Parve, Merike. 2000. On the acoustics of short monophthongs in Vastseliina (South Estonian). *Linguistica Uralica* 36(4). 241–260.
- Parve, Merike. 2003. Setu välted. *Lõunaeesti häälikud II* (Tartu Ülikooli Eesti Keele Õppetooli Toimetised 27). 17–47.
- Pihelgas, Kaire. 1999. Negationsform des Imperativs ohne k im Võru-Dialekt des Südestnischen. *Linguistica Uralica* 35(4). 255–268.
- Pihelgas, Kaire. 2000. *nud*-partitsiibi liitumine vokaaltüvele Võru murrakus. *Inter dialectos nominaque. Pühendusteos Mari Mustale* (Eesti Keele Instituudi Toimetised 7). 257–267.
- Pilvik, Maarja-Liisa. 2016. *Olema* + *Vmine*-konstruktsioonid eesti murretes. *Keel ja Kirjandus* 6. 429–446. <https://doi.org/10.54013/kk703a2>.
- Pilvik, Maarja-Liisa. 2017. Deverbal *-mine* action nominals in the Estonian Dialect Corpus. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 8(2). 295–326. <https://doi.org/10.12697/jeful.2017.8.2.10>.
- Pilvik, Maarja-Liisa. 2019. Assessing the productivity of the Estonian deverbal suffix *-mine* in five registers of Estonian. *SKY Journal of Linguistics* 32. 75–103.
- Pilvik, Maarja-Liisa. 2021. Comparing the productivity of Estonian deverbal suffixes *-mine*, *-us*, and *-ja* in five registers: A quantitative usage-based approach. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 12(1). 189–258. <https://doi.org/10.12697/jeful.2021.12.1.06>.
- Pilvik, Maarja-Liisa, Helen Plado & Liina Lindström. 2021. Murded, varieerumine ja korpusandmed. Eitussõna paiknemine võru ja seto eituslausetes. *Keel ja Kirjandus* 8–9. 771–796. <https://doi.org/10.54013/kk764a7>.
- Plado, Helen. 2015. *des-* ja *mata-*konverbi kasutusest eesti murretes. *Emakeele Seltsi Aastaraamat* 60(1). 195–218. <https://doi.org/10.3176/esa60.10>.
- Plado, Helen & Maria Reile. 2020. Vähemuskeele uurimine keeleteadusliku katsega: Demonstratiivpronoomenite kasutus ruumis viitamisel tänapäeva võru keeles. *Emakeele Seltsi aastaraamat* 65(65). 142–163. <https://doi.org/10.3176/esa65.06>.
- Pook, Hanna. 2018. *Pronoomeni kes kasutusest eesti murretes*. Tartu: University of Tartu. Master's Thesis.
- Pool, Raili. 1999. About the use of different forms of the first and second person singular personal pronouns in Estonian cases. *Estonian: Typological Studies III* (Tartu Ülikooli Eesti Keele Õppetooli Toimetised 11). 158–184.
- Prillop, Külli & Eva Saar. 2020. Eesti keele ajalooline morfoloogia. In Helle Metslang & Eva Velsker (eds.), *Eesti keele ajalugu* (Eesti Keele Varamu VI), 169–336. Tartu: Tartu Ülikooli Kirjastus.

- Proos, Mariann. 2020. Usage patterns of Estonian experience perception verbs: A random forest approach. *SKY Journal of Linguistics* 33. 67–101.
- Pung, Lea. 1968. *Tartu-Maarja murrak*. Tartu: University of Tartu. Diploma Thesis. <http://hdl.handle.net/10062/40487>.
- Põld, Enel. 2019. *Lääne-Saaremaa palatalisatsioonisüsteemi tänapäevane seis*. Tartu: University of Tartu. Master's Thesis. <http://hdl.handle.net/10062/64302>.
- R Core Team. 2018. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Reile, Maria. 2015. Space and demonstratives: An experiment with Estonian exophoric demonstratives. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 6(2). 137–165. <https://doi.org/10.12697/jeful.2015.6.2.06>.
- Reile, Maria. 2016. Distance, visual salience, and contrast expressed through different demonstrative systems: An experimental study in Estonian. *SKY Journal of Linguistics* 29. 63–94.
- Reile, Maria, Kristiina Averin & Nele Põldver. 2022. Kas see on sama mis too? Hinnangukatse eesti keele demonstratiivpronoomenite ja demonstratiivadverbidega. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 13(2). <https://doi.org/10.12697/jeful.2022.13.2.07>.
- Reile, Maria, Helen Plado, Harmen B. Gudde & Kenny R. Coventry. 2020. Demonstratives as spatial deictics or something more? Evidence from Common Estonian and Võro. *Folia Linguistica* 54(1). 167–195. <https://doi.org/10.1515/flin-2020-2030>.
- Reile, Maria, Piia Taremaa, Tiina Nahkola & Renate Pajusalu. 2019. Reference in the borderline of space and discourse: a free production experiment in Estonian, Finnish and Russian. *Linguistica Uralica* 55(3). 185–208. <https://doi.org/10.3176/lu.2019.3.02>.
- Rosenbach, Anette. 2013. Combining elicitation data with corpus data. In Manfred Krug & Julia Schlüter (eds.), *Research methods in language variation and change*, 278–294. New York: Cambridge University Press.
- Rosenberg, Taimi. 2000. Hargla murraku genitiiv. *Inter dialectos nominaque. Pühendus-teos Mari Mustale* (Eesti Keele Instituudi Toimetised 7). 320–339.
- Rosenberg, Taimi. 2002. Hargla murraku kohakäänded. *Nime murre. Pühendusteos Valdek Palli 75. sünnipäevaks* (Eesti Keele Instituudi Toimetised 11). 216–230.
- Ruutma, Mirjam. 2016. *Kaassõnad eesti murretes*. Tartu: University of Tartu. Master's Thesis. <http://hdl.handle.net/10062/51736>.
- Ruutma, Mirjam. 2019. Do contact languages influence the distribution of prepositions in Estonian dialects? *Nordic Journal of Linguistics* 42(3). 365–384. <https://doi.org/10.1017/S033258651900009X>.
- Ruutma, Mirjam, Aki-Juhani Kyröläinen, Maarja-Liisa Pilvik & Kristel Uiboaed. 2016. Ambipositsioonide morfosüntaktilise varieerumise kirjeldusi kvantitatiivsete profiilide abil. *Keel ja Kirjandus* 2. 92–113. <https://doi.org/10.54013/kk699a2>.
- Rätsep, Huno. 1979. *Eesti keele ajalooline morfoloogia II*. Tartu: TRÜ eesti keele kateeder.
- Saareste, Albert. 1933. Meie keele korraldamise põhimõttest ja nende kohaldamisest. *Eesti Kirjandus* 9. 446–454.
- Saareste, Andrus. 1924. *Leksikaalseist vahekordadest eesti murretes: I. Analüüs 60 kaardi ja I skeemiga*. Tartu: Acta et Commentationes Universitatis Dorpatensis.
- Saareste, Andrus. 1938. *Eesti murdeatlas. I vihk*. Tartu: Eesti Kirjanduse Selts.
- Saareste, Andrus. 1941. *Eesti murdeatlas. II vihk*. Tartu: Eesti Kirjanduse Selts.
- Saareste, Andrus. 1955. *Väike eesti murdeatlas*. Uppsala: Almqvist & Wiksell.

- Saareste, Andrus. 1958–1968. *Eesti keele mõisteline sõnaraamat I-IV*. Stockholm: Vaba Eesti.
- Sahkai, Heete. 2003. Demonstrative doubling in spoken Estonian. *Trames* 7(2). 120–144. <https://doi.org/10.3176/tr.2003.2.04>.
- Sang, Joel. 1974. *Eitavad laused Kihnu murrakus*. Tartu: University of Tartu. Diploma Thesis.
- Sang, Joel. 1975. Eitus Kihnu murrakus. *Keel ja Kirjandus* 3. 155–162.
- Sang, Joel. 1983. *Eitus eesti keeles*. Valgus.
- Schneider, Edgar W. 2008. Investigating Variation and Change in Written Documents. In Jack K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The Handbook of Language Variation and Change* (Blackwell Handbooks in Linguistics), 67–96. Malden: Blackwell.
- Sedrik, Meeli. 1994. Hiiu murrakute ja soome murrete ühissõnavarast. *Lähivõrdlusi. Lähivertailuja* 7. 103–127.
- Séguy, Jean. 1971. La relation entre la distance spatiale et la distance lexicale. *Revue de Linguistique Romane* 35(138). 335–357.
- Séguy, Jean. 1973. La dialectométrie dans l'Atlas linguistique de la Gascogne. *Revue de Linguistique Romane* 37(145). 1–24.
- Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Robert M. W. Dixon (ed.), *Grammatical Categories in Australian Languages*, 112–171. Canberra: Australian National University.
- Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin & Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC bioinformatics* 9(1). 1–11. <https://doi.org/10.1186/1471-2105-9-307>.
- Strobl, Carolin, James Malley & Gerhard Tutz. 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological methods* 14(4). 323–348. <https://doi.org/10.1037/a0016973>.
- Swart, Peter de, Monique Lamers & Sander Lestrade. 2008. Animacy, argument structure, and argument encoding. *Lingua* 118(2). 131–140. <https://doi.org/10.1016/j.lingua.2007.02.009>.
- Szmrecsanyi, Benedikt. 2013. *Grammatical variation in British English dialects: A study in corpus-based dialectometry* (Studies in English Language). Cambridge: Cambridge University Press.
- Szmrecsanyi, Benedikt. 2017. Variationist sociolinguistics and corpus-based variationist linguistics: overlap and cross-pollination potential. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 62(4). 685–701. <https://doi.org/10.1017/cnj.2017.34>.
- Szmrecsanyi, Benedikt & Lieselotte Anderwald. 2018. Corpus-based approaches to dialect study. In Charles Boberg, John Nerbonne & Dominic Watt (eds.), *The Handbook of Dialectology* (Blackwell Handbooks in Linguistics), 300–313. Oxford: John Wiley & Sons.
- Söderman, Tiina. 1996. *Lexical characteristics of the Estonian North Eastern coastal dialect*. Uppsala: University of Uppsala. PhD Thesis.
- Tagliamonte, Sali. 2012. *Variationist sociolinguistics: change, observation, interpretation* (Language in Society 40). Malden: Wiley-Blackwell.
- Tagliamonte, Sali A. 2013. Analysing and interpreting variation in the sociolinguistic tradition. In Manfred Krug & Julia Schlüter (eds.), *Research methods in language variation and change*, 382–401. New York: Cambridge University Press.

- Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language variation and change* 24(2). 135–178. <https://doi.org/10.1017/S0954394512000129>.
- Tanning, Salme. 1961. *Mulgi murdetekstid* (Eesti Murded I). Tallinn: Eesti Riiklik Kirjastus.
- Tanning, Salme. 1965. Kontraktaverbidest eesti murretes. *Emakeele Seltsi Aastaraamat* 11. 157–173.
- Tanning, Salme. 2004. *Karksi murrak*. Tallinn: Eesti Keele Sihtasutus.
- Taremaa, Piia. 2017. *Attention meets language: A corpus study on the expression of motion in Estonian*. Tartu: University of Tartu. PhD Thesis. <https://dspace.ut.ee/handle/10062/55362>.
- Tauli, Valter. 1966. *Structural tendencies in Uralic languages*. London/The Hague: Mouton.
- Tekko, Aino. 1958. *Vändra murraku monograafia*. Tartu: University of Tartu. Diploma Thesis. <http://hdl.handle.net/10062/28843>.
- Teras, Pire. 1998. About the acoustics of long and overlong vowels in the Võru dialect. *Linguistica Uralica* 34(3). 247.
- Teras, Pire. 2003. Võru keskkõrgete vokaalide kvaliteedist. *Lõunaeesti häälikud II* (Tartu Ülikooli Eesti Keele Õppetooli Toimetised 27). 48–67.
- Teras, Pire. 2007. Leivu hääldusjoontest. *Tartu Ülikooli Lõuna-Eesti keele- ja kultuuri-uuringute keskuse aastaraamat* VI. 19–30.
- Tirkkonen, Mari-Epp. 2007. Ranniku- ja kirdemurde endofoorsed pronoomenid. *Keel ja Kirjandus* 4. 277–290.
- Toomsalu, Ülo. 1995. *Setu verbi grammatika ja sõnastikud*. Tallinn: Perioodika.
- Tragel, Ilona & Jane Klavan. 2021. Kuhu suundub püsivus? Verbidega jääma ja jätma väljendatud sündmuste kujutamine katses. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 12(1). 369–395. <https://doi.org/10.12697/jeful.2021.12.1.10>.
- Türk, Helen, Eva Liina Asu, Pärtel Lippus & Ellen Niit. 2016. Diftongidest ja triftongidest eesti keeles Kihnu vokaalisüsteemi näitel. *Keel ja Kirjandus* 1. 1–12. <https://doi.org/10.54013/kk698a1>.
- Uiboaed, Kristel. 2010a. Ühendverbid eesti murretes. *Keel ja Kirjandus* 1. 17–36.
- Uiboaed, Kristel. 2010b. Statistilised meetodid murdekorpuse ühendverbide tuvastamisel. *Eesti Rakenduslingvistika Ühingu aastaraamat* (6). 307–326. <https://doi.org/10.5128/ERYa6.19>.
- Uiboaed, Kristel. 2013. *Verbiühendid eesti murretes*. Tartu: University of Tartu. PhD Thesis. <http://hdl.handle.net/10062/34499>.
- Uiboaed, Kristel, Cornelius Hasselblatt, Liina Lindström, Kadri Muischnek & John Nerbonne. 2013. Variation of verbal constructions in Estonian dialects. *Literary and linguistic computing* 28(1). 42–62. <https://doi.org/10.1093/lilc/fqs053>.
- Uiboaed, Kristel & Liina Lindström. 2014. Murrete lauseehitus ja selle uurimine. *Oma Keel* 1. 12–21.
- Vaba, Lembit. 1989a. Balti laenud läänemeresoome maastikusõnavaras. *Keel ja Kirjandus* 3. 138–141.
- Vaba, Lembit. 1989b. Balti laenud läänemeresoome maastikusõnavaras (järg). *Keel ja Kirjandus* 4. 206–218.
- Vaba, Lembit. 1997. *Uurimusi läti-eesti keelesuhetest*. Tallinn/Tampere: Eesti Keele Instituut.

- Vaba, Lembit. 2010. Kuidas läti-eeesti keelekontakt on mõjutanud eesti murdekeele grammatikat ja sõnamoodustust. *Emakeele Seltsi aastaraamat* (56). 204–246. <https://doi.org/10.3176/esa56.11>.
- Van Alsenoy, Lauren & Johan van der Auwera. 2015. Indefinite pronouns in Uralic languages. In Matti Miestamo, Anne Tamm & Beáta Wagner-Nagy (eds.), *Negation in Uralic languages*, 519–546. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Vangsnes, Øystein Alexander. 2007. Scandinavian dialect syntax (before and after) 2005. *Nordlyd* 34(1). <https://doi.org/10.7557/12.86>.
- Velsker, Eva. 2000. *Inessiivi lõpu varieerumine Vastseliina murrakus*. Tartu: University of Tartu. Master's Thesis.
- Veski, Johannes Voldemar. 1958. *Keelelisi töid*. Tallinn: Eesti Riiklik Kirjastus.
- Vihman, Virve-Anneli & Diane Nelson. 2019. Effects of animacy in grammar and cognition: Introduction to special issue. *Open Linguistics* 5(1). 260–267. <https://doi.org/10.1515/opli-2019-0015>.
- Viikberg, Jüri. 2020. *Eesti murrete grammatika* (Eesti Keele Varamu VIII). Tartu: Tartu Ülikooli Kirjastus.
- Viitso, Tiit-Rein & Valts Ernštreits. 2012. Līvōkīel-ēstikīel-leṭkīel sōnārōntōz. Liivieesti-lāti sōnaraamat. Lībiešu-igauņu-latviešu vārdnīca. Tartu and Rīga: Tartu Ülikool and Latviešu valodas aģentūra.
- Vija, Maigi. 2007. Pronoomenid lapsekeeles: mõnda *mina* ja *sina* omandamisest. *Eesti Rakenduslingvistika Ühingu aastaraamat* 3. 373–384. <https://doi.org/10.5128/ERYa3.24>.
- Viks, Ülle. 1973. Isevārki sõna. *Keel ja Kirjandus* 7. 421–427.
- Vilbaste, Gustav. 1993. *Eesti taimenimetused* (Emakeele Seltsi Toimetised 20). Tallinn: Eesti Teaduste Akadeemia, Emakeele Selts.
- Walker, David A. & Thomas J. Smith. 2020. Logistic regression under sparse data conditions. *Journal of Modern Applied Statistical Methods* 18(2). 2–18. <https://doi.org/10.22237/jmasm/1604190660>.
- Walker, James. 2013. Variation analysis. In Robert J. Podesva & Devyani Sharma (eds.), *Research methods in linguistics*, 440–459. New York: Cambridge University Press.
- Whaley, Lindsay J. 1996. *Introduction to typology: the unity and diversity of language*. Thousand Oaks: Sage Publications.
- Wiedemann, Ferdinand Johann. 2011. *Eesti keele grammatika*. (Ed.) Ellen Niit. (Trans.) Heli Laanekask. Tallinn: Eesti Teaduste Akadeemia Emakeele Selts.
- Wiese, Heike & Horst J. Simon. 2002. Grammatical properties of pronouns and their representation. In Horst J. Simon & Heike Wiese (eds.), *Pronouns – Grammar and Representation* (Linguistics Today 52). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Winkler, Eberhard & Karl Pajusalu. 2018. *Salis-Livisch II. Grammatik und Wörterverzeichnis. Mit einem Anhang zu den salis-livischen Sprichwörtern*. Wiesbaden: Harrassowitz Verlag.
- Witzlack-Makarevich, Alena & Ilja A. Seržant. 2018. Differential argument marking: Patterns of variation. *Diachrony of differential argument marking* 19. 1–40.
- Õim, Haldur. 2001. Is there a folk theory of Self: the case of Estonian *ise* and *enese-enda*. *Papers in Estonian Cognitive Linguistics* (Tartu Ülikooli Eesti Keele Õppetooli Toimetised 2). 7–21.

- Yamamoto, Mutsumi. 1999. *Animacy and reference: A cognitive approach to corpus linguistics* (Studies in Language Companion Series 46). Amsterdam: John Benjamins Publishing Company.
- Yelland, Phillip M. 2010. An introduction to correspondence analysis. *The Mathematica Journal* 12(1). 86–109. <https://doi.org/10.3888/tmj.12-4>.

PUBLICATIONS

CURRICULUM VITAE

Name: Hanna Pook
Date of birth: 10.03.1994
Citizenship: Estonian
Telephone: +372 5196 2396
E-mail: hanna.pook@ut.ee

Education

2018–2023 University of Tartu, Doctoral studies, Estonian and Finno-Ugric Linguistics (Estonian language)
2016–2018 University of Tartu, Master’s studies, Estonian and Finno-Ugric Linguistics (Estonian language), *cum laude*
2013–2016 University of Tartu, Bachelor’s studies, major in Mathematical Statistics, minor in Estonian and Finno-Ugric Linguistics
2002–2013 Tallinn English College

Professional employment

2022–... Institute of the Estonian Language, lexicographer/junior researcher (1.00)
2020–2022 Institute of the Estonian Language, lexicographer/junior researcher (0.30)

Publications

Pook, Hanna; Lindström, Liina (2022). The use of the indefinite pronoun *keegi* ‘someone’ in Estonian dialects. *Nordic Journal of Linguistics*, 1–32.
Pook, Hanna (2021). Object case variation of the pronoun *mis* ‘what’ in spontaneous spoken Estonian and Estonian dialects. *ESUKA-JEFUL*, 12 (1), 259–301.
Saar, Eva; Hagu, Paul; Käsi, Inge; Leivo, Maeve; Pook, Hanna; Pajusalu, Karl (2020). *Seto eripäraste sõnade sõnaraamat*. Tartu-Tallinn: EKSA.
Pook, Hanna (2019). The pronoun *kes* ‘who’ and its referent’s animacy in Estonian dialects. *SKY Journal of Linguistics*, 32, 105–144.
Pook, Hanna (2018). *Kommunitseerima ja kommunikeerima*. *Oma Keel*, 36 (1), 27–34.

ELULOOKIRJELDUS

Nimi: Hanna Pook
Sünniaeg: 10.03.1994
Kodakondsus: eesti
Telefon: +372 5196 2396
E-post: hanna.pook@ut.ee

Haridus

2018–2023 Tartu Ülikool, doktoriõpe, eesti ja soome-ugri keeleteadus (eesti keel)
2016–2018 Tartu Ülikool, magistriõpe, eesti ja soome-ugri keeleteadus (eesti keel), *cum laude*
2013–2016 Tartu Ülikool, bakalaureuseõpe, peaeriala matemaatiline statistika, kõrvaleriala eesti ja soome-ugri keeleteadus
2002–2013 Tallinna Inglise Kolledž, hõbemedal

Töökogemus

2022–... Eesti Keele Instituut, leksikograaf-nooremteadur (1,00)
2020–2022 Eesti Keele Instituut, leksikograaf-nooremteadur (0,30)

Publikatsioonid

Pook, Hanna; Lindström, Liina (2022). The use of the indefinite pronoun *keegi* ‘someone’ in Estonian dialects. *Nordic Journal of Linguistics*, 1–32.
Pook, Hanna (2021). Object case variation of the pronoun *mis* ‘what’ in spontaneous spoken Estonian and Estonian dialects. *ESUKA-JEFUL*, 12 (1), 259–301.
Saar, Eva; Hagu, Paul; Käsi, Inge; Leivo, Maeve; Pook, Hanna; Pajusalu, Karl (2020). Seto eripäraste sõnade sõnaraamat. Tartu-Tallinn: EKSA.
Pook, Hanna (2019). The pronoun *kes* ‘who’ and its referent’s animacy in Estonian dialects. *SKY Journal of Linguistics*, 32, 105–144.
Pook, Hanna (2018). Kommunitseerima ja kommunikeerima. *Oma Keel*, 36 (1), 27–34.

DISSERTATIONES PHILOLOGIAE ESTONICAE UNIVERSITATIS TARTUENSIS

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

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