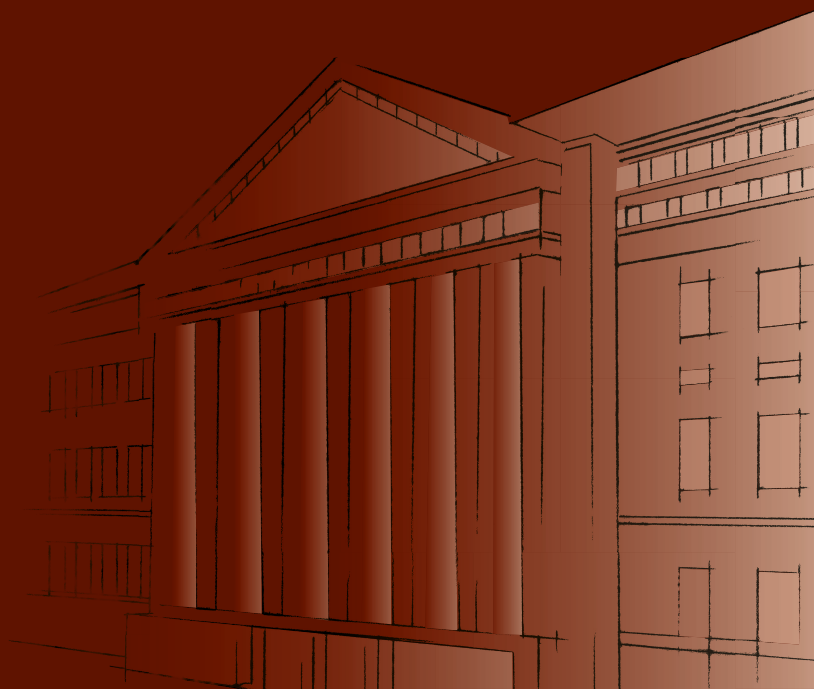


HELEN HINT

From full phrase to zero:  
a multifactorial, form-specific and  
crosslinguistic analysis  
of Estonian referential system





**HELEN HINT**

From full phrase to zero:  
a multifactorial, form-specific and  
crosslinguistic analysis  
of Estonian referential system



UNIVERSITY OF TARTU  
Press

University of Tartu, Institute of Estonian and General Linguistics

Dissertation accepted for the defence of the degree of Doctor of Philosophy on June 22<sup>nd</sup>, 2021 by the Committee of the Institute of Estonian and General Linguistics, Faculty of Arts and Humanities, University of Tartu

Supervisors: Prof Renate Pajusalu, PhD (University of Tartu)

Associate Prof Elsi Kaiser, PhD (University of Southern California)

Opponent: Prof Marja-Liisa Helasvuo, PhD (University of Turku)

Commencement: September 14<sup>th</sup>, 2021 at 14.15, in room 139, University Main Building, Ülikooli 18, Tartu, and in Zoom

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); and by the Estonian Research Council (PUT701).



European Union  
European Regional  
Development Fund



Investing  
in your future

ISSN 1406-5657

ISBN 978-9949-03-675-2 (print)

ISBN 978-9949-03-676-9 (pdf)

Copyright: Helen Hint, 2021

University of Tartu  
[www.tyk.ee](http://www.tyk.ee)

## ACKNOWLEDGEMENTS

The process of writing this dissertation officially started in Tartu in the year 2012. However, the initial spark of interest in the topic of referential devices was lit in me as early as in the autumn of 2008, when I attended the annual *teemalaat* (lit. ‘topic fair’) at our institute. At this event, I got excited about a possible BA thesis research topic that professor Renate Pajusalu introduced: *Referential features of noun phrases in Estonian*. I remember clearly why the study of reference and referential devices got my attention – it seemed to be the most difficult topic, compared to other possibilities presented at the *teemalaat*. (For those of you who wish to see a disclaimer here: yes, I might have been *a bit* naïve as undergrad with *a bit* of an exaggerated self-opinion back then.) In any way, so it happened that 13 years after this initial encounter, I am still obsessed with the topic of reference. It is only thanks to very many brilliant people who have, knowingly or unknowingly, guided, supported, and accompanied me on this extraordinary pilgrimage to unwrap this “most difficult” subject matter, that I withstood until the finale.

Before anybody else, I am overwhelmingly thankful to my supervisor Renate Pajusalu, who was keen to share her own research interests with me. As banal as it may sound, Renate really is the best supervisor. She was always there for me whenever I had questions or difficulties, to help me with her sharp and stimulating ideas and comments. I am especially thankful for her skill to unassumingly motivate me in those moments where it seemed easier to let go. Furthermore, Renate is a wonderful colleague and a good person, and it has been a pleasure and privilege to work with her.

I am also extremely grateful to my co-supervisor Elsi Kaiser, who kindly welcomed me at her psycholinguistics lab in Los Angeles. Although this thesis did not end up being a psycholinguistic study, Elsi’s psycholinguistic experiments and studies about Finnish and Estonian pronouns were the strongest source of inspiration for me when I was still writing up my PhD proposal. Elsi’s way of sharing her expertise enthusiastically, yet patiently, has helped me tremendously. Thank you for that!

The quality of this thesis has considerably increased thanks to thorough and thoughtful comments given by the three reviewers of this thesis, Marja-Liisa Helasvuo, Jorrig Vogels, and Virve Vihman. I greatly appreciate the time they took for reading the thesis. Their comments were not only necessary and useful, but also gave me a good opportunity to spend some more quality time with the topic I really enjoy thinking about.

This thesis would not be here without the great collaboration with my co-authors and project fellows. Thank you, Renate Pajusalu, Tiina Nahkola, Piia Taremaa, and Maria Reile, for writing and thinking with me. I greatly value the smart, inspiring, and honest teamwork we had, and will continue to have, I hope.

Throughout my studies at the University of Tartu, many good teachers and colleagues have supported my process of becoming an independent thinker. I

am deeply indebted to Ilona Tragel, who took me and my fellow students as equal colleagues at the time when we were just linguist babies, entering our first seminars in cognitive linguistics. Ilona's seminars always called for intellectual effort, but were at the same time held in a fun and relaxed atmosphere with the greatest company, a combination which I truly enjoyed. I also thank all my colleagues at Tartu, but especially Liina Lindström, Virve Vihman, and Ann Veismann, for their willingness to have longer or shorter discussions with me, either in seminars, in the corridors of the Jakobi building, or at conferences, on the topics that they know a lot more about than I do. These discussions have been very helpful for devising various pieces of this thesis.

For writing this thesis, I have received financial support from several projects and institutions, which have enabled me to carry out my research.

- Estonian Research Council grant “Referential devices in Estonian and adjacent languages: experimental approach” (PUT701);
- the project “Language and meaning: semantics and grammar in a cognitive perspective” (SF0180056s08);
- the Centre of Excellence in Estonian Studies (European Union, European Regional Development Fund);
- the European Social Fund's Doctoral Studies and Internationalisation Programme DoRa, which is carried out by the Archimedes Foundation;
- the Ministry of Education and Research, the Kristjan Jaak programme;
- the Graduate School of Linguistics, Philosophy and Semiotics at the University of Tartu.

Writing a thesis takes a lot of, well, writing. I was lucky to receive excellent writing guidance and assistance from two awe-inspiring colleagues and teachers, Anni Jürine and Djuddah Leijen, from the very beginning of my PhD studies. It is only thanks to you two that I learned how to really put together everything that is between these covers. I am also truly happy that I had a chance to participate in our lively and supporting writing group throughout these nine years. Maria Reile, Piia Taremaa, Mariann Proos, Nele Ots, Mari Aigro, Anni Jürine, Jane Klavan, Djuddah Leijen, Roger Yallop, and everybody else who has participated in the group at different times, your feedback and encouragement was vital for this thesis.

In relation to writing, I have to confess that producing texts in English is definitely not one of my greatest expertise. I am thankful to Katriin Hint, Uldis Balodis, Roger Yallop, Djuddah Leijen, and the editors at Keelekord, who helped to make my texts readable.

Anni and Djuddah, you have assured me that there will be life in academia, an exciting new life, after defending this thesis. I cannot put into words how fortunate I am to be part of the Bwrite project. This project is the reason why my PhD defence does not look like an end, but as a new beginning. Anni, I will always remember you as my great mentor and coolest colleague. You are a (rock)star, and I can always turn to your star when (writing) questions go over my head. Djuddah, I am so very grateful for your trust and support. Also, thanks

for cheering me up these very last agonizing weeks of the final revision. Your advice to let go came just when I needed it the most.

At the beginning of my studies in general linguistics at Tartu, I had the good fortune to make friends with Kaisa, Andres, and Martin. Your company through all these years have always delighted me. The next *kursapidu* is on me! My very special thanks goes to Pille for laughing and whining with me about the ups and downs of PhD life. Mariliis, Mirjam, Kärt, and Kadri, you have always been there for me, be it for a fun girls' night out with a glass of wine, for good advice on any urgent issue, or for burning discussions about anything not related to my thesis (that is, about politics, beauty and fashion, kids, men, or feminism). You have showed me the beauty of true friendship.

I am blessed to have a family who always has my back. Mom and dad, you have never questioned my decisions, and you have never stopped believing that what I do is right. The fact that you always take my boys under your wings when I need to do my “very important” work stuff tells it all. Marleen, it is quite something that we get exmatriculated (in a positive way) in the same year, you 10 years and I 15 years after our initial admission to university. I am so sorry that because of me being such a brainy bookworm, you have constantly had to listen to their nagging about how your grades should be better. I know that you know it, but still: you are the best version of you, and the best version of a little sister, regardless of the stupid grades.

And finally, I get to thank the ones who mean the most to me. When I started writing this thesis, I had no idea that there could be anything more awesome in the world than having a PhD degree. Today I know that you my dear boys, Mihkel and Erik, are thousands of times more awesome than whichever degree. You are the greatest achievements of my life. Meelis, when we first met in the spring of 2008, I was just a second-year undergrad, but I confidently proclaimed that I aim to pursue a career in academia, which, as is known, requires writing a PhD thesis. Call it faith, a coincidence, or just magic of numbers, but we got married in 2012. You have patiently put up with my stubborn goals for all these years. I wish that my time in academia would last as long as our walk together – until the end of time.

# CONTENTS

LIST OF ORIGINAL PUBLICATIONS .....	10
ABBREVIATIONS.....	11
INTRODUCTION.....	13
1. THEORETICAL UNDERPINNINGS AND MAIN OBJECTIVES OF THE THESIS.....	16
2. BASIC CONCEPTS AND APPROACHES IN STUDIES OF REFERENCE.....	20
2.1. The essence of reference.....	20
2.1.1. What is reference? .....	20
2.1.2. Anaphoric and deictic reference.....	22
2.2. Noun phrases as referential devices.....	23
2.2.1. Full vs. reduced referential devices .....	25
2.2.2. Definite vs. indefinite referential devices.....	28
2.3. Factors that affect referential choice .....	29
2.3.1. Monofactorial salience-based accounts of reference.....	29
2.3.2. Multifactorial approach to reference .....	32
3. ANAPHORIC REFERENTIAL DEVICES IN ESTONIAN.....	34
3.1. A summary of previous studies and the niche of this thesis.....	34
3.2. Grammatical features that shape the Estonian referential system .....	37
3.3. Referential devices in Estonian .....	40
3.3.1. Estonian article-like determiners .....	42
3.3.2. Estonian demonstratives as independent devices .....	43
3.3.3. Estonian third person personal pronouns.....	45
3.3.4. Zero reference in Estonian.....	46
3.3.5. The relationship between different types of NPs.....	48
4. METHODOLOGY.....	50
4.1. Data collection methods .....	50
4.1.1. Spoken narrative data .....	51
4.1.2. Written corpus data.....	54
4.2. Description of explanatory variables.....	55
4.3. Statistical methods used for analysing the data .....	59
4.3.1. Descriptive statistics.....	59
4.3.2. Trees and forests.....	60
5. MAIN RESULTS AND DISCUSSION.....	61
5.1. The system of Estonian referential devices .....	61
5.1.1. Zero reference.....	61
5.1.2. Article-like determiners.....	64
5.1.3. Anaphoric pronouns .....	67



5.1.4. Demonstratives .....	70
5.1.5. Outline of the Estonian referential system .....	72
5.2. A cross-linguistic perspective on referential devices .....	76
6. METHODOLOGICAL IMPLICATIONS AND LIMITATIONS OF DATA COLLECTION AND ANALYSIS .....	80
6.1. Collecting data for finding referential devices in discourse .....	80
6.1.1. Narrative data .....	80
6.1.2. Corpus data .....	82
6.2. Coding data for analysing referential devices .....	84
6.2.1. Detecting and tagging for possible predictor variables .....	84
6.2.2. Finding the alternative variants .....	86
7. CONCLUSION .....	88
REFERENCES .....	92
SUMMARY IN ESTONIAN .....	109
PUBLICATIONS .....	125
CURRICULUM VITAE .....	264
ELULOOKIRJELDUS .....	266

## LIST OF ORIGINAL PUBLICATIONS

This dissertation is based on four original publications. Throughout the dissertation, these studies will be referred to with the following names in small caps.

**ZERO:** 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.

**DETERMINERS:** Hint, Helen, Tiina Nahkola and Renate Pajusalu. 2017. With or without articles? A comparison of article-like determiners in Estonian and Finnish. *Lähivõrdlusi. Lähivertailuja* 27. 65–106.

**PRONOUNS:** Hint, Helen, Tiina Nahkola and Renate Pajusalu. 2020. Pronouns as referential devices in Estonian, Finnish, and Russian. *Journal of Pragmatics* 155. 43–63.

**DEMONSTRATIVES:** Hint, Helen, Piia Taremaa, Maria Reile and Renate Pajusalu. 2021. Demonstratiivpronoomenid ja -adverbid määratlejatena: miks me oleme siin ilmas, selles olukorras? [Demonstrative pronouns and demonstrative adverbs as determiners in Estonian: why are we in “here world” in “this situation”?]. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics* (in press).

### Contribution to the publications

**ZERO:** Helen Hint was the sole author of the article.

**DETERMINERS:** Helen Hint designed the data collection method and was the main deviser of the data coding schema. She collected, transcribed, and coded the Estonian data and double-coded the Finnish data. She was responsible for interpreting the results and running the statistical analysis. 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.

**PRONOUNS:** Helen Hint designed the data collection method and was the main deviser of the data coding schema (coinciding with DETERMINERS). She collected, transcribed, and coded the Estonian data and double-coded the Finnish and Russian data. She was responsible for interpreting the results and running the statistical analysis. 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.

**DEMONSTRATIVES:** Helen Hint initiated the research. She collaborated with other authors to devise the method and the coding schema. She tagged the 2400 corpus clauses for the variable ‘Semantic class’ and conducted the ‘tree and forest’ analysis of the data. She was the corresponding author of the article and was responsible for correcting and revising the manuscript.

## ABBREVIATIONS

### List of abbreviations used in the glosses

ADE	adessive
ALL	allative
COM	comitative
COND	conditional
DEM	demonstrative pronoun
ELA	elative
GEN	genitive
INE	inessive
INF	infinitive
LOC	locative
LONG	long form of personal pronoun
NOM	nominative
Ø	zero reference
PL	plural
PRS	present
PRT	partitive
PST	past
SG	singular
SHORT	short form of personal pronoun

### List of other abbreviations used in the text

adj	adjective
adv	adverb
Cb	backward-looking centre
CT	Centering theory
def	definite
dem	demonstrative
demAdv	demonstrative adverb
demPron	demonstrative pronoun
det	determiner
detNP	NP with a determiner
fullNP	full noun phrase
genAtt	genitive attribute
genNP	NP with a genitive attribute
indef	indefinite
N	noun
NP	noun phrase
OtherAnim	other animate entities in a clause
otherNP	NP with other determiner
persPron	personal pronoun
PrevRefNP	previous referential NP

pron	pronoun
RefDist	referential distance
<i>see</i> NP	NP with a determiner <i>see</i> ‘this/the’
SubstSem	semantic class of substantive
SyntRole	syntactic role
<i>üks</i> NP	NP with a determiner <i>üks</i> ‘a/one’

## INTRODUCTION

**Background and importance of the study.** In verbal communication, people need to point out certain events, processes or states, and specify who or what performs these actions and to whom/what these actions are directed. On the level of speech acts, this is described as a distinction between acts of predication and acts of reference (Searle 1969; P. Hanks 2019). The persons and objects involved in the events, usually called entities, can be either animate (*a teacher*) or inanimate (*a hat*), concrete (*the teacher*, *the hat*) or abstract (*goodness*), and they can function as the main arguments of the verb (e.g., *a teacher writes*) or only as some clausal components with lesser importance (e.g., *the teacher with a red hat writes*). Furthermore, the entities can exist in the real world (e.g., *our English teacher*) or they can exist only in the discourse world (e.g., *Harry Potter's potions teacher*). Either way, the interacting speakers typically pay a great deal of attention to the entities that are the basis for the mental representations. The process of mentioning the entities is captured by the term 'reference'.

Reference is a multifaceted process. On one hand, there are far fewer words and expressions in a language than there are entities in the world. For example, *him* can, in principle, refer to any male person, actual or fictional, who lives or has ever lived in the universe. On the other hand, a wide array of linguistic expressions may be used for referring to one specific entity. For example, a man can be referred to with many expressions, like *the man with a hat*, *a smart guy*, *this man*, *the person*, or *he* (the list can be continued, of course). Furthermore, a referential act may be grounded in a purely textual content (i.e., anaphoric reference), or it may retrieve necessary context from the surrounding physical space (i.e., deictic reference). How, then, does a speaker decide which expression to use for reference in a particular discourse? This has been a primary question since the early studies about reference (e.g., Bühler 1990 [1934]; Chafe 1976; Givón 1983a; Gundel et al. 1993).

To understand the underlying principles of referential choice, it must be taken into account that reference is a discourse process and referents are discourse elements that are created and change in the course of interaction (Cornish 1999; Etelämäki 2009; W. F. Hanks 1990). In a stretch of natural text, referential devices referring to specific entities usually vary: a speaker might start a story by introducing a referent, *an old lady*, then mention this referent again as *the woman* and continue talking about this referent as *she*. However, when the speaker introduces another female character into the same story, say *a smart girl*, it is probably the case that the pronoun *she* is used later in the text to refer either to the old lady or to the smart girl. Communication is effective only when a relatively similar mental representation of the discourse is shared by the participants in the conversation. In addition, the use of referential devices is not only affected by the ongoing context, but referential choice itself functions as a means for creating context (Silverstein 1976; Laury 1996a).

Traditionally, the choice of a referential device in a discourse is believed to reflect the mental salience of the referents in that discourse (Givón 1983a; Ariel 1990; Gundel, Hedberg, and Zacharski 1993). According to such salience-based approaches, more reduced devices (e.g., pronouns) are used for referring to more salient referents, whereas less salient referents are mentioned with more elaborate forms (e.g., full noun phrases). However, later studies have effectively shown that a single factor-based approach to referential choice is not sufficient. Instead, several factors, possibly at the same time, are at play when the speaker chooses a referential expression (Kaiser and Trueswell 2008; Kibrik 2011). This thesis follows the idea that referential choice is related to several factors in addition to salience (e.g., syntactic role, case, animacy, and referential distance). Furthermore, the factors included here represent different levels of language structure: namely, the effect of syntactic, semantic and pragmatic variables on referential choice.

This thesis is first and foremost an explorative study about the phenomenon of anaphoric reference in Estonian. Under the term ‘anaphoric referential devices’, several phrase types are included, such as personal and demonstrative pronouns, zero reference and full noun phrases with or without a determiner. In the past, Estonian referential devices have typically been studied from the perspective of their position in the grammatical system (Erelt et al. 1993; Pajusalu 2005, 2009). Also, the usage patterns of individual devices have been studied and explained in terms of their linguistic functions (e.g., Pajusalu 1997c, 2000, 2006a). This study builds on those previous works but takes a different perspective, so that the focus is not on individual devices. Instead, the thesis aims to determine i) how different referential devices work together in a complete discourse, and ii) which referential devices are used and how they alternate in referring to different discourse entities throughout the discourse. In addition, the Estonian referential system is compared to corresponding patterns in Finnish and Russian. The comparison is necessary for assessing whether and how the use of seemingly similar referential devices across languages is affected by various factors. Also, this thesis addresses the ways that referential systems differ across languages due to language-specific properties.

For studying referential choice, natural discourse data is most often deemed to be the best. On that account, this thesis focuses on quasi-experimentally elicited spoken narrative discourse to reveal the common usage patterns of referential devices. In addition, written corpus data, representing natural texts of different types and purposes, was collected for ascertaining the behavioural properties of certain devices which do not appear in narrative discourse as often. Besides its potential for explaining concrete aspects of particular referential devices, corpus data exhibits promising features for the study of referential devices in general. A unifying aspect throughout this thesis, however, is the multifactorial analysis of the data, for which conditional random forests and inference trees are employed. Furthermore, in the individual studies collected for this thesis, qualitative examples and explanations of the data are emphasized.

The results of the thesis contribute to previous research on reference in two ways. Firstly, it is the first comprehensive, form-specific multiple-constraint approach to Estonian referential devices. The study lists all major types of referential devices in Estonian and clarifies their referential properties individually. Consequently, the traditional salience-based approach to referential choice is expanded, and it is shown that besides salience, many other factors guide the choice of a proper referential expression. Furthermore, it is shown that referential choice in Estonian is dependent mostly on pragmatic factors in the case of NPs with determiners, whereas the use of pronouns is more closely related to grammatical factors. This is argued to be a language-specific feature since different patterns emerge in other languages. Secondly, this work refines some methodological procedures for studying referential devices. A straightforward method is demonstrated for researchers who wish to strictly compare the usage patterns of similar devices across languages or the patterns of referring to certain kinds of entities in different languages.

**The structure of the thesis.** The thesis is structured as follows. In **Chapter 1**, the theoretical foundations, aims and objectives of the thesis, both theoretical and methodological, are listed. **Chapter 2** elaborates on the theoretical background of the study, explains the notion of reference and related concepts, and outlines different views on the phenomenon of reference. **Chapter 3** focuses on the system of Estonian referential devices. The chapter starts with an overview of the study of reference in Estonian and proceeds toward a systematic description of Estonian referential expressions, including reduced anaphoric devices as well as referential full NPs. **Chapter 4** is concerned with methodological issues. The material elicitation methods and procedures, as well as analysis techniques are explained. **Chapter 5** presents the main results of the study and discusses these findings, including both Estonian-specific as well as crosslinguistic outcomes and methodological aspects. In **Chapter 6**, the methodological implications and limitations that relate to data collection and data coding processes are presented. The chapter discusses some practical considerations for conducting the research reported in this thesis, and possible prospects for planning future studies. The thesis concludes with **Chapter 7**, which summarises the most important outcomes and implications of the thesis.

# 1. THEORETICAL UNDERPINNINGS AND MAIN OBJECTIVES OF THE THESIS

This thesis draws upon three different, yet interrelated, perspectives. First, it is an explorative research study about the general system of referential devices in Estonian. Second, the thesis is designed to offer an explanation for the not yet fully understood differences between referential devices that have supposedly similar functions in Estonian. Third, this study includes a cross-linguistic inspection of the systems of referential devices in three geographically close languages, Estonian, Finnish, and Russian, that share some important formal features but also have significant individual variance in the use of referential devices.

The overall theoretical basis of this thesis lies in the cognitive-functional school of research. More specifically, the thesis is framed by the theoretical foundations of cognitive linguistics, which connects language use and the human cognitive system. The founding fathers of cognitive linguistics, Langacker (1987) and Lakoff (1987, 1990), have formulated the idea that grammar is an integral part of the human cognitive system and, as such, grammatical structures are meaningful units that help us conceptualise and engage the world. However, cognitive linguistics has traditionally shown limited interest in natural discourse data and discourse-related phenomena (for some noteworthy exceptions, see Van Hoek 1997; Goldberg (ed.) 2006).

Cognitive linguistics belongs to a larger group of functionalist approaches to language (e.g., Givón 1995; Kibrik and Plungian 1997; Nuyts 2007). These approaches place communication and information processing at the focus of interest and suggest that language should be viewed and analysed from the perspective of its functional purposes, not merely from its formal component parts. The study of reference, thus, is well situated under this functionalist approach, since referential processes have first and foremost a discourse-based nature in language, and are cognitively determined. In this thesis, the more specific functionalist approach adhered to is the form-specific multiple constraints concept to reference, as introduced by Kaiser and Trueswell (2008) and Kibrik (2011). These approaches to referential choice emphasize that the salience-driven one-factor based account alone is not sufficient enough to exhaustively explain the reasons why one or another referential form is chosen in a particular context (Brown-Schmidt, Byron, and Tanenhaus 2005; Kibrik 1996; 2011; Kaiser and Trueswell 2008).

While the main theoretical underpinnings remain consistent throughout the thesis, every individual article in this thesis builds on different theoretical account. In *ZERO*, Centering theory (CT) is used (Grosz et al. 1995). In *DETERMINERS*, the discussion is built on the grammaticalization approach (Heine and Kuteva 2006). In *PRONOUNS*, the form-specific multiple constraint approach to reference is taken as a main starting point (Kaiser and Trueswell 2008; Kibrik 2011). The fourth study, *DEMONSTRATIVES*, follows the ideas of



Construction Grammar (Goldberg 1995, 2006) and alternation studies (e.g., Arppe 2009; Gries 2017). All these approaches are used as tools for solving specific tasks with a slightly different aim. However, each task seeks to explain certain functional aspects of the use of different referential devices and are, thus, complementary with each other. The different theoretical approaches that are used in this thesis are chosen so that they can be used to put together a multifaceted picture of the Estonian referential system.

Centering theory (Grosz et al. 1995; Walker et al. 1998), a starting point of the first study ZERO, stands theoretically most further away from the other studies in the thesis. CT is a framework specifically designed to explain discourse coherence within a discourse segment by focussing on the use and interpretation of referential devices in natural discourse (Gundel 1998: 190). As the theory operates with cognitive concepts such as ‘inference load’ and ‘attentional state’, it is clear that CT has common ground with other cognitive-functional approaches. What separates CT from other cognitive approaches, however, is its computational foundation with a rigid system of rules and constraints that allows us to account for a particular level of coherence in a particular segment. In the context of this thesis, CT is considered to be a useful linguistic tool with solid explanatory power, and as such, it allows us to investigate what kind of effect different referential forms (overt third person pronoun and zero reference in this study) have on the overall coherence of a discourse segment. While CT has lost much its influence over time, and while it is concluded in this thesis that CT alone does not have enough strength to fully determine the referential properties of various devices, it still remains relevant for researchers who are interested in the precise description of aspects of coherence and its underlying causes (see also Taboada and Wiesemann 2010). The narrow scope of the framework (i.e., the ambition to resolve only issues related to the use of referential devices and coherence) conveniently allow to pair the framework with other, more general frameworks.

Grammaticalization (Heine and Kuteva 2006) is another useful means for observing the place and function of certain linguistic devices. It is especially useful when these devices are newer parts of a language’s grammatical system, or even are just under way of becoming the full-fledged members of this system. In DETERMINERS, grammaticalization proves to be useful for the discussion about whether article-like determiners (e.g., *see* ‘this’ and *üks* ‘one/a’ in Estonian) should be treated as devices that bear the tasks of grammatical articles, or are they rather pragmatic units that are used under certain discourse conditions.

Goldberg’s (1995, 2006) Construction Grammar, as another approach situated in cognitive linguistics, is used in DEMONSTRATIVES. The goal of the study DEMONSTRATIVES was to apply Construction Grammar and alternation studies framework (e.g., Arppe 2009; Janda and Solovyev 2009; Gries 2017) to explain the semantic difference between two seemingly similar determiner constructions: one with a demonstrative adverb as a determiner of the NP, and the other with a demonstrative pronoun as a determiner of the NP. Construction Grammar was chosen for this purpose because it provides the means for explaining dif-

ferences between certain referential devices, analysed as constructions in this case, that have similar referential properties in terms of salience and linguistic structure, but that nevertheless have different functions. Alternation studies as a methodological approach (Gries 2017) also complement the form-specific multiple constraints approach, since it allows to take into account the influence of several factors at once.

The chosen theoretical frameworks often operate with different basic terms and concepts. For example, CT presupposes the adoption of ‘centers’, ‘transitions’, and ‘focus of attention’, and Construction Grammar has ‘construction’ as an elementary unit. Some of the terms are framework specific and their usage do not extend over framework boundaries. Nevertheless, as all the theoretical approaches used in this thesis belong to the cognitive-functional school of thought, they do not contradict or eliminate each other. The most important theoretical concepts and the more exact definitions are explained in more detail in Chapter 2.

A thorough analysis of Estonian (and Finnish and Russian) referential devices offers new insights into the field of reference studies that is currently shaped by the structure of English, the most often studied (Indo-European) language. Estonian and its contact languages, however, have many features that differ from the English system, for example, the lack of grammatical articles, the elaborate case system, and relatively free word order. The role of these features in referential choice, and their impact on the overall referential system, have not yet been comprehensively discussed. Therefore, this thesis also broadens the typological knowledge about the functioning of referential systems in different languages.

The more detailed objectives of the thesis are as follows:

- 1) Overall, to investigate which referential devices emerge as the core of the Estonian referential system, i.e., which devices most typically form the coherent referential chains in discourse (investigated in the studies ZERO, DETERMINERS, PRONOUNS, DEMONSTRATIVES).
- 2) To uncover the referent-related, meaning-related and language structure-related factors beyond salience that enable us to give a more comprehensive explanation of referential choice in discourse (studies ZERO, DETERMINERS, PRONOUNS, DEMONSTRATIVES).
- 3) To explore the distinctions between the referential properties of certain pairs of referential devices that are subject to constructional variation in Estonian. More specifically, the thesis studies i) the distinction between the overt third person pronoun *ta* vs. zero reference (ZERO), ii) the differences between demonstrative pronouns and demonstrative adverbs when used as a determiner in a noun phrase (DEMONSTRATIVES), iii) NPs with determiners vs. NPs without determiners (DETERMINERS), and iv) pronouns vs. fullNPs (PRONOUNS).

4) The thesis also aims to add a cross-linguistic perspective to this research by addressing the question of what kind of similarities and differences are exhibited in usage patterns of similar referential devices across languages. Namely, i) definite and indefinite determiners in Estonian and Finnish are contrasted (DETERMINERS), and ii) different pronominal forms and the underlying factors which affect the usage patterns of these forms in Estonian, Finnish, and Russian are compared (PRONOUNS).

In addition to these theoretical objectives, the thesis also confronts the following methodological issues:

1) The thesis aims to establish a straightforward data collection method for crosslinguistic comparison. Namely, to compare truly similar referential devices across languages, i.e., to pinpoint the similar referents in discourse and particular referential devices used for referring to them, a picture sequence-based narrative elicitation method is tested (DETERMINERS and PRONOUNS).

2) Reference can be most effectively studied as a discourse phenomenon (see also Kibrik 2011), i.e., the more context is available for the researcher, the better. In agreement with this assumption, this thesis employs full narratives as data in three studies. However, corpus data is used in the fourth study, DEMONSTRATIVES, which focuses on the referential properties of two determiner constructions with similar form and structure (det + NP), differentiated only by the determiner form. The purpose is to demonstrate that corpus data, usually appearing with considerably less context, is also suitable for studying referential properties of various devices.

3) In relation to the previous point, this thesis also seeks a method for investigating the traits of such seemingly similar forms in which i) one is clearly visible and overt, whereas the other is a covert zero expression, or ii) one is frequent in a certain data type (spoken narratives), but the other occurs there much less often (DEMONSTRATIVES).

4) Last but not least, this research tests the tools offered by Centering theory as an analytical and explanatory framework for differentiating between particular referential forms. The thesis also tests how compatible this framework is with Estonian data (study ZERO).

## 2. BASIC CONCEPTS AND APPROACHES IN STUDIES OF REFERENCE

The chapter begins by defining and explaining the concept of reference, as it is understood here. In addition, the main theoretical approaches to reference are described in this chapter. In the first section of this chapter (2.1), the fundamental concepts related to reference in linguistics are explained. The second section (2.2) elaborates on the different referential forms and expressions that speakers use for referring. Finally, the third part of this chapter (2.3) reviews previous research that examines which factors affect the choice of a referential expression and how.

### 2.1. The essence of reference

#### 2.1.1. What is reference?

The notion of reference has been of interest in many disciplines, such as philosophy, linguistics, psychology, etc. Due to its cross-disciplinary nature, it is not a simple task to exactly define ‘reference’, as different disciplines focus on different senses of reference. For a transparent explanation, it is first useful to distinguish between semantic reference and pragmatic reference (Abbott 2010; see also Gundel and Abbott 2019). **Semantic** reference is understood as a phenomenon related purely to linguistic expressions. In this tradition, it is assumed that linguistic expressions themselves have a connection to things in the world. The semantic sense of reference is more related to views relevant to philosophers, e.g., Frege, Wittgenstein, Russell, and Carnap.

The **pragmatic** interpretation of reference, on the other hand, emphasises the use of language by the speaker. As such, the pragmatic sense is more prevalent among linguists and states that reference is a three-place relation between the speaker, the linguistic device and the entity referred to (Abbott 2010: 2). This thesis is written in line with linguistic approaches and thus deals with pragmatic reference.

Studying pragmatic reference in natural language use requires that the researcher must also account for the interactional essence of the phenomenon. This means that not only the speaker, but also the addressee(s) must be included in the three-place description of reference. This is because the successful use of referential expressions in discourse requires that the speaker account for the knowledge and mental states of the interlocutor as well (Ariel 1990, 2001; W. F. Hanks 1990; Gundel, Hedberg, and Zacharski 1993; Gundel et al. 2010; Chafe 1994; Lambrecht 1994; Kibrik 2011). The underlying principles of how our cognitive processes and mental states guide the choice of a referential expression are concisely presented by Kibrik (2011, Ch. 10).

An entity can be any physical or mental construct (e.g., person, animal, object, or abstract phenomenon) that is perceived as relatively stable in time and that can be referred to with a linguistic device (Lyons 1977: 442–445). If an entity is evoked with a linguistic device in a particular clause, then it becomes the referent of that expression (Lambrecht 1994: 37). For example, *he*, *my new laptop*, *John's anger*, and *wind* all make equally good entities and referents. However, the physical existence of the entity in the text-external world is not a prerequisite for a referential act to be successfully performed. Therefore, it is useful to think about referents as conceptualizations in our minds (Jackendoff 2002; Vogels, Krahmer, and Maes 2019). After all, people frequently talk about things that do not really exist: consider, for example, fairy tales or fiction in general. It is perfectly fine to talk about Albus Dumbledore as the Headmaster of the Hogwarts School of Witchcraft and Wizardry and be totally aware that the existence of the person and place is beyond the evidence-based world view. This is only possible when both the speaker and the addressee have a (somewhat similar) mental representation of the referent in their heads (Chafe 1994; Laury 2001; Abbott 2010).

Referents are discourse elements that are created and modified in the course of interaction (W. F. Hanks 1990; Cornish 1999; Etelämäki 2009). When the speaker has a specific entity, either text-external or text-internal, in her mind, she can produce an explicit linguistic utterance that includes a potentially referential expression. If the addressee can pick up this referential expression and its intended referent, then the utterance can be considered referential. For example, when the speaker says, 'The classmate gave Mary a present because it was her birthday', she intuitively assumes that the addressee perceives the expressions *the classmate*, *Mary* and *her* as referential, and based on the given context, can relatively easily identify which real-world entities are talked about or create new referents. Furthermore, the addressee should be able to understand that *her* is just another way for referring to the already-mentioned *Mary*. If the addressee cannot retrieve the referent of the word *her* from the discourse, then the referential act has not been completed to its full potential (Chafe 1994; Hanks 1990; Lambrecht 1994). As such, reference should be considered as a part of larger phenomenon of text coherence (Halliday and Hasan 1976).

As a matter of communication, the act of referring normally involves more than one participant, i.e., at least one speaker and one addressee. One of the most important considerations in choosing a referential device is understandability: the speaker should make a choice based on assumptions about which devices are the clearest and most easily accessible to the addressee (see e.g., Ariel 1990; Chafe 1994; Gundel et al. 1993, 2010; Kehler et al. 2008; Kibrik 2011). For that reason, the analysis of referential expressions is inevitably connected to real interaction, and many researchers agree on the importance of a mutually negotiated referential framework, based on the joint attention of interlocutors (Diessel 2006; W. F. Hanks 1992; Laury 1997; Helasvuo 2020).

Acknowledging the influence of context on the choice and usage patterns of referential devices is fundamental for understanding this discourse phenomenon. Crucially, the speech context has two parties: the speaker and the addressee. The

speaker chooses and produces the referential form, but the addressee must interpret it by drawing a connection between an expression and a referent. Thus, borrowing from Cornish (1999: 20): “the establishment of a referent is a joint achievement, undertaken by the interlocutors collaboratively, and is not simply the responsibility of the speaker alone”. Furthermore, more recent studies even propose three-party models in which story characters should also be included in dynamic networks of communication, at least when considering narrative discourse (Verhagen 2019; Dancygier 2019). According to this view, story characters have similar mental and communicative abilities as the speaker and addressee, and the viewpoint of the character is another aspect in the text that influences the choice of deictic expressions.

### 2.1.2. Anaphoric and deictic reference

Whether the speaker relies on the text-internal or text-external world has further important implications for the phenomenon of referential choice. Traditionally, referential acts have been divided into **deictic** (exophoric, spatial) and **anaphoric** (endophoric, textual) based on the contextual factors of a speech setting (Fillmore 1997; Cornish 1999; Halliday and Hasan 1976; Kibrik 2011; Lambrecht 1994). A referential act is deictic if the referent has been identified using the physical context of the interlocutors. A referential act is anaphoric if the referent is identified based on the previous conversation or text. When an interlocutor exclaims ‘Look at this!’ and points her finger toward a giant spider, this is a classic example of deictic reference, since the entity referred to (*the spider*) is physically present in the actual surroundings. However, when your friend tells you about her dreadful meeting with a giant spider the day before and says, ‘There was an enormous spider in my bathroom. It wanted to kill me’, she relies on the ongoing text to establish the expression’s referent. Thereafter, she can refer to the spider anaphorically (*it*).

The distinction between deictic and anaphoric reference originates with Bühler’s (1990 [1934]) fundamental treatment of deictic expressions as the spatio-temporal co-ordinates of a speech event that relate to the *origo* as the zero-point of the deictic centre. Bühler divided deictic expressions into three categories: i) reference in the immediate surroundings of the interlocutors, ii) text-internal reference, and iii) imagination-oriented deixis. Depending on the nature of the referent, it is customary to distinguish between i) spatial deixis, ii) time deixis, and iii) person deixis (*Ibid.*). In Bühler’s sense, deictic expressions also include anaphoric reference (text-internal reference). However, there are also approaches that make a stricter division between these two. For example, Fillmore (1997) lists five categories of deixis (person deixis, place deixis, time deixis, discourse deixis and social deixis), and none of these include anaphoric, text-internal devices.

Several authors, however, suggest that deictic and anaphoric reference are better captured by a more holistic view and that the two uses should be analysed

together (Laury 1997; Etelämäki 2009; Jarbou 2010; Talmy 2017). For example, Talmy (2017) argues that it is not possible to strictly distinguish between the two types, speech-external and speech-internal targets, in his treatment, since the basic mechanisms ('targeting') by which the reference is conveyed are the same in both cases. Talmy (2017: 3–13) specifies as much as ten cue categories (e.g., lexical cue categories, bodily cue categories, background cue categories, etc.) that help to direct the hearer's attention to a particular 'target' (a notion that, in principle, corresponds to the 'referent'), and these cue categories have the potential to interact and supplement one another (*Ibid.*: 13–21).

Moreover, and following the ideas proposed by Talmy (2017), the relationship between a referential device and space goes beyond mere anaphoric-deictic distinction. Namely, the conceptualization of a referent can depend on the choice of a referential device, so that certain aspects of the referent are focused on or specified by using a certain device, no matter whether they are anaphoric or deictic (see also Faure 2014). Also, when linking a word to the world, the use of referential devices (especially demonstratives) is not only dependent on the context, but the devices are themselves a means of creating context (Silverstein 1976: 34). For example, Laury (1996, 1997) has suggested that in Finnish, the choice of a demonstrative expression and its grammatical form does not always depend on the properties of the referent but is rather a way of conceptualizing the referent. According to Laury (1997: 145), locations conceptualized as figures tend to be expressed with demonstratives, while ground-like conceptualizations turn up as adverbials. Therefore, in the case of spatial relations, as discussed in DEMONSTRATIVES, a referent can be conceptualized as space or as an entity, depending on which determiner is chosen as an NP attribute. Namely, it is possible to use a demonstrative pronoun as a determiner in Estonian, e.g., *selles majas* 'in this house', but a demonstrative adverb can also fill a determiner position, e.g., *siin majas* 'in this house [here]'. Both constructions can be used in anaphoric as well as in deictic context.

Therefore, although this thesis employs the term 'anaphoric reference' for the sake of clarity, it does not imply an intrinsic divergence between anaphoric and deictic reference. The devices investigated in this study can be used for deictic reference as well. This means that the distinction between anaphoric and deictic reference does not lie in the specific referential forms, but in the usage contexts and purposes of these forms.

## 2.2. Noun phrases as referential devices

It is claimed that all languages make a distinction between reference and predication, or on the level of lexical classes, between nouns and verbs (e.g., Whaley 1997: 59). Therefore, a speaker is involved in (at least) two kinds of propositional acts when producing a sentence: a referential act (expressing animate entities/objects/places/abstract concepts) and a predicating act related to the referent (expressing actions/events/states) (Searle 1969). In terms of linguistic

structure, referential acts are transmitted by using different kinds of referential devices. As presumed by the pragmatic conception of reference, only nominal expressions can be viewed as grammatically manifesting referential functions in language (Abbott 2010; Kibrik 2011). Thus, specific reference<sup>1</sup> is a phenomenon usually expressed with noun phrases (NPs) in language (Helasvuo 2020). The objective of this section is to illustrate different types of referential NPs and to elaborate the fundamental difference between full and reduced referential devices.

The types of basic referential devices, as understood in the thesis, follow the overall systematisation in Kibrik (2011), but individual definitions are specified, where necessary. ‘Full noun phrase’ stands for an informative NP that specifies and categorises the referent lexico-semantically. The main subtypes of full NPs are proper names (*Mary*) and common nouns. The latter, in turn, can be divided into bare nouns (*kindness, a car*), NPs with a determiner (*this book, his car*), or NPs with short or long modifier (*a blue car, the book I read yesterday*) (Kibrik 2011: 38). ‘Pronouns’ are reduced but overt referential devices that can be used to substitute full NPs in discourse (Crystal 2008; Kibrik 2011: 39). There are several classes of pronouns, but from the perspective of reference, personal and demonstrative pronouns are the most important. ‘Personal pronouns’ are connected to the category of person and they are used for either deictic (first and second person, i.e., *I* and *you*) or anaphoric (third person, e.g., *he*) reference (Siewierska 2004). Demonstratives are primarily understood as deictic expressions (Bühler 1990 [1934]; Levinson 2006). Nevertheless, ‘demonstrative pronouns’ (*this, that*) very often have anaphoric uses, and they are used to mention activated referents in discourse, similarly to personal pronouns (Himmelman 1996; Levinson 2006; Kibrik 2011: 124). Furthermore, ‘demonstrative adverbs’ are seen here as separate type of referential devices that denote locative referents (e.g., *here, there*) (Laury 1996a; Nahkola et al. 2020).

Besides overt anaphoric reference, speakers can use ‘zero reference’, which is understood in this thesis as the “omission” of an overt referential form, i.e., the referent is implicitly mentioned in a clause, but no overt formal device appears (e.g., *Peter ate ice cream, then  $\emptyset$  drank coffee*) (Kibrik 2011: 104–105). In the literature, the phenomenon of a “missing” anaphoric referential form has been marked with a plethora of terms, for example referential zero (Kibrik 1996), zero person form (Siewierska 2004), zero or zero pronominal (Ariel 1990; Gundel et al. 1993), zero anaphora (Obana 2003; Oh 2006; Tao and Healy 2005), null pronoun (Fedele and Kaiser 2014; Filiaci, Sorace, and Carreiras 2014; Gelormini-Lezama and Almor 2011), null argument (Takahashi 2008), and null subject (Haegeman 2007; Schmitz, Di Venzio, and Scherger 2016; Scott 2013).<sup>2</sup>

---

<sup>1</sup> There are also instances of non-specific and non-referential uses of noun phrases. These include, for example, generic, attributive and predicative reference (see Kibrik 2011: 32). Such uses are not dealt with in the present thesis.

<sup>2</sup> Some related terms are also (argument) ellipsis (Duguine 2014; Lindström 2013; Takahashi 2008), (subject) omission (Haegeman 2013; Lindström and Vihman 2017), and pro-drop (Duguine 2014; Kwon and Sturt 2013; Weir 2012).



These different terms are often not clearly or unequivocally defined, and it can also happen that the same authors find different terms useful in different occasions. The terms that explicitly encompass the concept of pronoun/anaphora mostly refer to the formally unfilled argument or adjunct positions that correspond to particular referents (Kibrik 1996: 261). A more rigid position is taken by Siewierska (2004: 22–23), who applies the term zero in absolute sense, meaning that there exists no phonological form at all for a grammatical person interpretation, and, according to this view, even person inflection solely on the verb should be treated as means of person reference. In this thesis, the term ‘zero reference’ is the preferred one, since it implies the realisation of a referential act, but it does not make any assumptions about the grammatical nature of the device.

The following passage (1) illustrates the use of different types of referential NPs.

- (1) *Betsey gave the small girl a pet hamster as a present. She was shy when passing it to the girl.*

The underlined expressions all refer to some entity. Co-referential expressions, i.e., expressions that refer to the same entity, are indicated by different types of underlining. It can easily be noted that the referential expressions differ a lot in their length and semantic content. In (1), four different NP types are used: proper noun (*Betsey*), indefinite NP with an article (*a pet hamster*), definite NP with an article (*the small girl*, *the girl*), and personal pronoun (*she*, *it*). The possible reasons the speaker chooses a particular type at a certain point in the discourse is the essential question in the field of reference. To answer this question, a list of all possible referential NP types must be compiled beforehand.

### 2.2.1. Full vs. reduced referential devices

A noun phrase is, by definition, expected to contain a noun. However, there are certain subcategories of NPs, for example, pronouns, that go beyond that delimitation (Abbott 2010: 4). Independent pronouns (personal and demonstrative) can form an NP on their own, or they can function as determiners and modify head nouns in an NP (Helasvuo 2020). Therefore, in classifying different types of referential devices, the most fundamental distinction is made between full NPs vs. pronouns, i.e., between full vs. reduced<sup>3</sup> referential devices (Kibrik 2011). Full NPs have enough lexical-semantic content for the referent to be understood in isolation (Pajusalu 2017a; Karimi et al. 2014). On the opposite

---

<sup>3</sup> Instead of the term ‘reduced referential device’, other concepts have also been used, for example, ‘attenuated referential device’ (Chafe 1994), ‘minimal reference’ (Laury 2005), or ‘semantically poorer form’ (Ariel 1994). ‘Reduced referential devices’ should not be confused with ‘reduced pronominals’. The latter term has been used by Siewierska (2004) and denotes pronominal functions with no overt form, as opposed to ‘full pronominals’.

side are reduced referential devices, which are informationally impoverished (Kaiser and Trueswell 2008), i.e., very vague in their lexical-semantic content and thus cannot specify the referent when isolated from the larger context (Garnham 2001: 55; Kibrik 2011: 38–39). In this thesis, thus, reduction is defined in terms of the device’s semantic capacity, i.e., how much semantic knowledge can be obtained from the device itself, when not taking the surrounding context into account. In that sense, full referential devices are proper names and full NPs with or without modifiers. Various pronouns and zero reference are reduced referential devices.

The types of possible referential expressions have been arranged into different lists with varying degrees of precision by different authors (Givón 1983b; Ariel 1990; 2001; Gundel, Hedberg, and Zacharski 1993). Often, the underlying assumption characteristic to such lists is that a referential device is chosen based on the salience/givenness/mental accessibility of the expression/referent: more reduced devices are used for referring to more salient referents (this topic will be discussed in more detail in section 2.3.1). Probably the most exhaustive list of different possible referential NPs is given by Ariel (1990: 73), whose scale of accessibility contains almost 20 different types of referential expressions. From that list, it can be inferred that there is considerable structural variation in the linguistic expression of reference. The Givenness Hierarchy by Gundel et al. (1993), with ‘only’ six statuses, is less complex at first glance, but that does not reduce the complexity of the linguistic system, since one cognitive status may be expressed by more than one linguistic device.

The main types of referential devices that emerge from literature are: i) full NPs without a determiner (e.g., *maja* ‘house’, *vesi* ‘water’), ii) full NPs with a definite or indefinite determiner (e.g., *üks maja* ‘a/one house’, *see maja* ‘the/this house’), iii) demonstratives (e.g., *see* ‘this’, *too* ‘that’, *siin* ‘here’), iv) stressed personal pronouns (e.g., *tema* ‘s/he’), v) unstressed personal pronouns (e.g., *ta* ‘s/he’), and vi) zero reference (i.e., *ø söi putru* ‘ø ate oatmeal’).<sup>4</sup>

Researchers dealing with various aspects of reference have been perhaps most interested in the usage patterns of reduced referential expressions, i.e., different types of pronouns, such as personal pronouns vs. demonstrative pronouns (e.g., Kaiser and Trueswell 2008; Kaiser and Vihman 2010; Kaiser 2011b), overt vs. zero reference (e.g., Ueno and Kehler 2016; Schmitz, Di Venzio, and Scherger 2016; Scott 2013; Pekelis 2018), proximal vs. distal demonstratives (e.g., Etelämäki 2009; Coventry, Griffiths, and Hamilton 2014; Rocca, Tylén, and Wallentin 2019; Reile 2019; Reile et al. 2019), and full NPs vs. pronouns (e.g., Gordon, Grosz, and Gilliom 1993; Arnold and Zerkle 2019; Arnold and Griffin 2007; Vogels, Krahmer, and Maes 2012; 2015).

---

<sup>4</sup> Ariel (1990) has included verbal person inflections in her list of referential expressions as well. See also Siewierska (2004) for a discussion about the referential properties of verbal inflection. Although it is agreed here that verbal inflections have certain referential properties, this is not perceived as their primary function. Thus, verbal person inflections are not discussed as referential devices in this thesis.

In terms of **personal pronouns**, only third person forms (e.g., *he, she*; Estonian *ta/tema*) are relevant to anaphoric reference (Chafe 1994; Siewierska 2004). First and second person reference cannot be straightforwardly compared to third person, because first and second person are deictic speech-act pronouns in nature; i.e., their referents are always present in the speech situation as the speaker and the addressee (Siewierska 2004; Lindström et al. 2008; Metslang 2013). Third person reference, in contrast, is anaphoric, meaning that its referent can be identified based on the ongoing discourse context and must be previously mentioned with an explicit NP (Siewierska 2004: 7).<sup>5</sup>

**Demonstratives** (e.g., *this, that*; Estonian *see, too*) are very often analysed in terms of deictic, situational use (Diessel 1999, 2006; Fillmore 1997; Coventry et al. 2008), but demonstratives are common as anaphoric referential devices as well (Himmelman 1996; Levinson 2006; Gundel et al. 1993; Ariel 1990). As discussed above (section 2.1.2), it has been suggested that it is not useful to make a sharp distinction between deictic and anaphoric uses of demonstratives, since the deictic field that shapes the use of demonstratives in interaction is dynamic (Etelämäki 2009; W. F. Hanks 1992; Laury 1997; Talmy 2017). In terms of anaphoric reference, demonstratives are often considered as evoking less salient referents than pronouns. However, as will be discussed in section 2.3.2, this assumption may not be valid cross-linguistically.

**Zero reference** is identified in this thesis as the “omission” of an overt referential form (Kibrik 2011: 104–105). However, some terminological discrepancy can be found in study ZERO, where the following terms are used irregularly: zero person marker, zero, zero reference, zero pronominal, zero pronoun, and zero pronominal form. All these terms are to be understood as synonyms without differences in their meaning. For the sake of clarity, however, the use of synonymous terms is avoided in later studies of this thesis.

Here, it is important to distinguish between the so-called zero-person construction and zero reference. Zero person constructions have implicit arguments and a generic reading, meaning that the agent is not specified in such constructions (Kaiser and Vihman 2006). In the syntactic description of Estonian, such constructions are referred to as ‘defective clauses’ (Erelt 2017: 87–88). An example of zero-person construction is presented in (2), where it is not possible to deduce any overt agent (i.e., person), although the finite verb takes the third person singular form. Thus, although the sentence specifies the possible action (to *swim*), it does not give a specific doer, because the sentence is meant to be about ‘anyone’, i.e., non-referential.

- (2) *Suvel*                      *saab*                      *seal*                      *ujuda*.  
 Summer.ADE              get.PRS.3SG              there              swim.INF  
 ‘In summer, you/one can swim there.’

<sup>5</sup> Some exceptions that contradict the requirement of previous mention can appear in natural language use, see for example Greene et al. (1994) about unheralded pronouns that lack explicit antecedent in the text.

Zero reference, on the other hand, is clearly referential and expresses a particular referent in an ongoing discourse. Zero reference is used, for example, in the passage presented in (3). Here, the finite verb is also in the third person singular form, and although the agent is not expressed with an overt phonological form (except the verbal inflection), the identity of the agent (*laps* ‘child’) is clear based on the overall context and preceding referential expressions. In this thesis, only instances of zero reference for referring to a clause subject are analysed.

- (3)    *laps*                    *sõi*                    *järve*                    *ääres*                    *jäätist*  
          child.NOM        eat.PST.3SG        lake.GEN        by                    ice cream.PRT  
          *siis*                    *läks*                     $\emptyset$                     *ujuma*  
          then                    go.PST.3SG         $\emptyset$                     swim.INF  
          ‘the child ate an ice cream by the lake. then  $\emptyset$  went for a swim’

### 2.2.2. Definite vs. indefinite referential devices

As indicated above, a distinction is made between definite and indefinite NPs. Definiteness usually relates to the identifiability<sup>6</sup> of the referent, and it relates to the speaker’s expectations about the attentional states of the addressee and informational status of the referent (Chafe 1994; Chesterman 1991; Du Bois 1980; Laury 2001; C. Lyons 1999). The category of definiteness is often associated with formal aspects of referential devices, especially in languages with grammatical articles. In article languages, the indefiniteness of an NP is usually signalled by an indefinite article, and the definite article is an indicator of a definite NP (Chesterman 1991; C. Lyons 1999; Laury 2001; Dryer 2013b; 2013a; Abbott 2006).

Even languages that lack proper articles can sometimes use determiners to explicitly express the informational status of a referent. Determiners are defined as a group of functional words (including articles, demonstratives, interrogative determiners, etc.) that can operate as an attribute of the head noun in an NP, and as such, express the informational properties of the referent (Payne 2006; Pajusalu 2017a). Thus, grammatical articles are usually covered with the term of determiner, along with devices such as adnominally occurring demonstratives, indefinite pronouns, possessives, and other pronouns. Other form-related options for marking definiteness in non-article languages also exist; for example, in Estonian and Finnish, the case form of syntactic argument can express certain aspects of definiteness (Chesterman 1991; Metslang 2012). The relationship between case forms and definiteness, however, is not discussed further here.

<sup>6</sup> Different notions have been used in literature, e.g., activation (Givón 1993; Kibrik 1996), givenness and consciousness (Chafe 1994; see also Gundel et al. 1993), accessibility (Ariel 2001), and salience (Kaiser and Trueswell 2008). Although these notions are not entirely synonymous, they are all concerned with the cognitive (attentional) aspects of the locutors.

Still, definiteness cannot be formally marked on every referential NPs, even in article-languages. For example, pronouns and proper names (usually) do not occur with articles or with determiners overall. However, personal pronouns, demonstratives, proper nouns, and NPs with possessives, along with NPs that exhibit explicit definiteness marking, are considered to be definite NPs (Abbott 2006; Laury 2001). Furthermore, the source of definiteness<sup>7</sup> can have an effect on the choice and interpretation of the referential form (Fraurud 1990; C. Lyons 1999). For example, a referent most often becomes definite via previous mention. However, in certain contexts, a speaker may use a definite expression for referring to an entity that has not been mentioned before in an ongoing discourse (e.g., *the Sun* is always considered definite). Therefore, while first-mentioned referents are very often indefinite, they may also appear as definite under certain conditions.

Therefore, we see that definiteness is not fully guided by grammatical, but also by pragmatic principles. Previous research that has addressed this issue tends to make a distinction between a formal category of definiteness and a pragmatic/semantic category of identifiability (Du Bois 1980a; Prince 1992; Chafe 1994; Laury 2001). Importantly, the identifiability of a referent (and the definiteness of an NP) can only be deduced in the ongoing discourse. Thus, there exist NPs whose definiteness cannot be explicitly stated based on their linguistic form, since they can be used for definite as well as for indefinite reference, depending on their usage context (Kibrik 2011; Laury 2001).

## 2.3. Factors that affect referential choice

This section summarises two different approaches to referential choice by focusing on the factors that determine it. The section begins with an overview of a group of approaches that adopt salience as the one guiding factor (section 2.3.1). It continues by outlining later approaches that expand the initial salience-based view and instead support the multifactorial concept of reference (section 2.3.2).

### 2.3.1. Monofactorial salience-based accounts of reference

Textual coherence, and successful reference as means of creating coherence (Halliday and Hasan 1976), is not an objective quality of the text, but it relates to the mental processes involved in discourse production (Givón 1993). Con-

---

<sup>7</sup> An NP may become definite through various sources. First, the discourse itself (anaphoric vs. deictic) is influential. In the linguistic context, there may be different ways of becoming definite, e.g., previous mention, an earlier mention of a related referent (e.g., *bike*; *front wheel*), the uniqueness of the referent (e.g., *the Queen of England*), etc. See Fraurud (1990), Chesterman (1991) and Lyons (1999) for further detail.

sequently, pragmatic approaches to reference usually take a cognition-based approach to explain referential choice. Attentional states and the location of the referent in memory are most often seen as the aspects responsible for referential choice (Gundel et al. 1993). Activation of a referent in memory is directly related to focus of attention (Kibrik 1996, 2011). More elaborate forms (e.g., *a black poodle*) relate to referents that are mentioned for the first time in a discourse, i.e., are not yet in the focus of attention, and as such, are not salient. Shorter reduced forms (e.g., *that* or *he*), on the other hand, refer to already-mentioned, i.e., salient entities in the focus of attention (Chafe 1994).

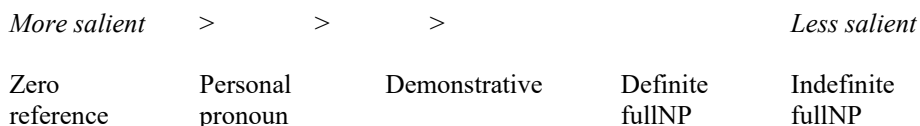
As a result, the ground category that is seen to lie behind the referential choice is salience (also accessibility, givenness, activation status, prominence)<sup>8</sup> (Givón 1983a; Ariel 1990; Gundel, Hedberg, and Zacharski 1993; Chafe 1994). In this thesis, ‘salience’ (understood as a synonym to ‘accessibility’) is used as an umbrella term to refer to the qualities of information that make it easy for the interlocutors to cognitively access this information (Arnold 2010; Vogels, Krahmer, and Maes 2019). By using this term, it is not intended to make any assumptions about the specific cognitive domains that are responsible for establishing the information as salient. Nevertheless, the term covers unquestionably the interconnection of salience and cognitive processes.

Based on the correlation between a referential form and the salience of the referent, several hierarchies have been proposed, of which the most influential are probably Gundel, Hedberg, and Zacharski (1993), Givón (1983a) and Ariel (1990). The term ‘pragmatic selection models’ has been used by Arnold and Zerkle (2019) to describe the models of referential choice presented in this section. The main tenets of one specific pragmatic selection model, namely Givenness Hierarchy proposed by Gundel et al. (1993), are as follows: i) the form of a referring expression depends on the assumed cognitive status of the referent, ii) there is a certain number of cognitive statuses (six in Givenness Hierarchy) that relate to different forms of referential expression used in discourse, and iii) by hearing a particular form, the addressee can infer the corresponding referent and its cognitive status. For example, when a speaker uses a definite article, she indicates that the referent should be uniquely identifiable to the addressee, and when she uses the pronoun *it*, the speaker indicates that the referent holds the highest possible cognitive status.

Gundel et al. (1993) claim that the hierarchy is universal and the Givenness Hierarchy is suitable for explaining referential choice in every language, with the reservation that all statuses are not necessarily required in every language. It is of course not unexpected that languages differ in their linguistic inventories. Therefore, more standard hierarchies, based on summarising various pragmatic selection models, have been proposed, such as the one presented in Figure 1, where less salient forms appear to the right, and more salient forms to the left (Kaiser and Trueswell 2008).

---

<sup>8</sup> See Arnold (2010) and Vogels, Krahmer, and Maes (2019) for discussions about the relationship between the terms.



**Figure 1.** The standard hierarchy of referential forms along the salience scale

Many factors have proven to influence the salience of the referent and hence the choice of a referential device: information structure (Colonna et al. 2012; Crawley 1986; Gernsbacher 1990; Järvisikivi et al. 2005), syntactic role (Crawley, Stevenson, and Kleinman 1990; Grosz, Joshi, and Weinstein 1995), grammatical role parallelism (Chambers and Smyth 1998; Sauermann and Gagarina 2017), thematic role of the referent (Arnold 2001; Stevenson, Crawley, and Kleinman 1994), semantics and general world knowledge (Hobbs 1979), and different coherence relations (Kehler et al. 2008). However, it is not clearly explained in salience-only accounts how all these factors relate to each other, and whether all these factors have equal effect on all different referential devices, or only on some of the devices.

A slightly different framework about referential choice is Centering theory, which is explained as a model of the speakers' center of attention in discourse that is concerned with the relationship of attentional state, inferential complexity, and the form of referring expression (Walker et al. 1998: 1; see also Grosz et al. 1995). By this definition, Centering is also a pragmatic selection model and hence shares similar underlying principles with other salience-based approaches (see Arnold and Zerkle 2019). However, Centering differs from the aforementioned approaches by a rigid system of rules and constraints, which are rooted in the computational base of Centering theory. Center, the basic term of the theory, is a semantic entity in an utterance that links this utterance to other utterances within a discourse segment. There are three kinds of centers: forward-looking, backward-looking, and preferred centers. Based on the comparison of the utterance's center and the center in the preceding utterance, the transition types are computed. Four types of transitions are postulated in Centering: CONTINUE, RETAIN, SMOOTH SHIFT, and ROUGH SHIFT. These transitions provide the means to explain how coherence is achieved (Walker, Joshi, and Prince 1998; Grosz, Joshi, and Weinstein 1995; Taboada and Hadic Zabala 2008). A more elaborate description of Centering and its application to Estonian data is presented in study ZERO in this thesis.

The main problem with the salience-only based approaches, as also pointed out by Kibrik (2011: 389–390), is the circularity of reasoning. Namely, the aforementioned approaches explain referential choice in terms of salience or similar cognitive status, and at the same time, the corresponding status is inferred from the particular linguistic device that is used for referring in particular linguistic context. Therefore, the linguistic device itself serves as the basis for deciding over the level of salience. For example, when the personal pronoun (e.g., *ta* 'he') is observed in the text, this leads to a conclusion that the referent of

this device is highly salient. Therefore, a more elaborate approach is necessary to overcome the problems that arise in mono-factorial models of reference.

### 2.3.2. Multifactorial approach to reference

Salience has remained one of the key concepts in explaining referential choice. Nevertheless, linguists and psycholinguists have presented findings that challenge the unified salience-based approach. For example, Kibrik (1996, see also 2011) has demonstrated, based on Russian written narrative discourse, that several factors (e.g., rhetorical distance, syntactic role, animacy, protagonist-hood) simultaneously affect the activation, and thus the pronominalisability, of a referent. Ariel (2001: 29) indicates that in English, *it* and *this/that* are not distinguishable based just on their informative content. Kaiser and Trueswell (2008) have shown that in Finnish, the personal pronoun *hän* ‘s/he’ is sensitive to the syntactic role of the preceding co-referential NP, but the choice of the demonstrative *tämä* ‘this’ relates to word order and syntactic role. Research by Scott (2013) demonstrates how Givenness Hierarchy (Gundel et al. 1993) alone is unable to offer an explanation for English overt and zero reference alternation, since both devices express the highest cognitive status (‘in focus’). Another problem, referring to the horizontal dimension of the Givenness Hierarchy, is that in Russian (and probably in other languages), an NP with an adnominal demonstrative (*ëtot*NP) can relate to any one of six statuses in the hierarchy (Krasavina 2011).

These studies reveal (at least) two problems with the salience-based approach. First, the use of different referential expressions might be triggered by different factors, which may, but do not have to, be related to salience. Second, different forms might express similar levels of salience, and thus, salience is not sufficient to explain these differences. As a result, more recent approaches to referential choice emphasize that the salience-based account alone is not enough to exhaustively explain the reasons why one or another form is chosen (Brown-Schmidt, Byron, and Tanenhaus 2005; Kibrik 1996, 2011; Kaiser and Trueswell 2008).

The number of possible factors that affect referential choice is considerable. Kibrik et al. (2016) have tested as many as 25 factors that have proven relevant to the choice of referring expression, including referent-related, anaphor-related, antecedent-related and distance-related factors. It is, however, impossible to present a complete list of factors that have proven useful in explaining referential choice. One reason is that different forms have been shown to react differently to various factors (Kaiser and Trueswell 2008). This means that the lists of influential factors might be different for different types of expressions. Secondly, the effect can vary depending on the internal structure of a language, genre and register, the individual speaker, etc. (Kibrik 1996, 2011).

Another approach that demonstrates how a single scale of pronoun interpretation is not sufficient to reveal all the workings of reference resolution is offered by Kehler and Rohde (2013). They propose a probabilistic Bayesian



model that aims to reconcile the coherence-based (Hobbs 1979) and Centering-based (Grosz, Joshi, and Weinstein 1995) approaches to reference. Kehler and Rohde’s model is appealing since it brings together the addressee’s expectations and the speaker’s referential choices. As the model’s cross-linguistic applicability has been questioned (Kaiser 2013), this line of research is not further developed in this thesis. However, Kehler and Rohde’s reconciliation of two seemingly diverse approaches to reference resolution is taken as a yet another indication that we need more elaborate ways to reveal how the referential system of language works than just one factor-based explanation.

The conception of referential choice as a form-specific multiple constraints process is central to the discussion in this thesis, as well, because it has several advantages over the single-factor approach. First and foremost, the multitude of factors enables to avoid the circularity in reasoning that arises in the salience-only account. Second, the multifactorial approach allows to bring together factors from various levels of linguistic description, some of which are often ignored in reference studies, e.g., case as a grammatical factor, or semantic class of the noun as a semantic factor. In addition, it does not challenge the approach to include both discourse-related factors (e.g., referential distance) and factors relating to referent’s internal properties (e.g., animacy) (see also Kibrik et al. 2016). Third, by acknowledging the simultaneous role of several factors, it is not assumed that some factors (e.g., salience) should be taken as principal to all other factors to subsume under. Instead, the interplay between all the factors indicates the salience level. However, in this thesis, it is assumed that salience does not depend on a fixated set of features, and the sets of influential variables are probably different across various referential devices.

It is necessary to consider the possibility that significant variation exists across languages in terms of how sensitive a particular referential device is to a particular factor (see also Kibrik 1996). Moreover, as suggested by Haspelmath (2010), devices which share similar grammatical labels (e.g., personal pronoun, demonstrative, determiner, etc.), might be used for very different functions. This is, of course, a more general problem pertaining to the grammatical description of languages. Studies DETERMINERS and PRONOUNS in this thesis set out to investigate more closely this possibility of factorial variation across various referential devices.

### **3. ANAPHORIC REFERENTIAL DEVICES IN ESTONIAN**

The following chapter describes in greater detail the studies related to various referential devices available in Estonian and referential choice between them. First, a systematic overview about the study of reference in Estonian is given in section 3.1. In order to better understand the overall organization of the Estonian system, section 3.2 presents a summary of the grammatical features relevant to the use of Estonian referential devices. After that, the results from the described studies are grouped and presented according to different referential devices in section 3.3. The overview proceeds along the line of salience of the described device(s), starting with devices that express the least salient entities in discourse. First, the functions and usage patterns of article-like determiners in Estonian are described (section 3.3.1). Estonian demonstrative pronouns as referential devices are then characterized in section 3.3.2. Section 3.3.3 is devoted to the system of Estonian personal pronouns, and previous studies that tackle the questions related to zero reference in Estonian are presented in section 3.3.4. The final section 3.3.5 summarises the chapter by showing the relationship between the described devices.

#### **3.1. A summary of previous studies and the niche of this thesis**

Before the 1990s, a decade that can be considered to be the start of research of Estonian referential devices, only some general observations about the system of Estonian noun phrases were presented. For example, some textbooks about Estonian grammar had presented the structure of noun phrases and described the usage contexts (and rules) of different pronouns (Tauli 1980; Mihkla et al. 1974). The knowledge collected on the topic was summarised in the academic treatment of Estonian grammar (Erelt et al. 1993); for example, the overall usage patterns of pronouns were described there. This grammar already mentions the referential functions of noun phrases, which are described under the term of substitution (*‘asendus’*) and, to some extent, also ellipsis (*‘väljajätt’*) (*Ibid.*).

The study of reference is closely connected to somewhat wider fields of information structure (e.g., Halliday 1985; Chafe 1976; Lambrecht 1994) and textual coherence and cohesion (Halliday and Hasan 1976). In the Estonian research context, however, the studies that tackle the essence of Estonian information structure focus instead on word order questions (Tael 1988; Lindström 2005). Text cohesion and coherence are more often used to explain the behaviour of Estonian NPs, for example, Erelt et al. (1993) describes the functions of Estonian NPs according to the notions of theme and rheme and also definiteness and indefiniteness. In addition, Tiit Hennoste (1986) has used the notion of cohesion to explain the principles of vocabulary choice in texts.

The more thorough investigation of reference in Estonian began in the 1990s with the work of Renate Pajusalu, who defended her dissertation “Deiktikud eesti keeles” (“Deictics in Estonian”) in 1999. At that time, investigations regarding referential devices in the Finnish language (e.g., Laury 1991, 1994, 1996a, 1996b; Seppänen 1998; Vilkuna 1992) served as an inspiration to Estonian linguists. In these early days of Estonian reference studies, the focus was on the general qualitative description of the Estonian pronoun system and certain referential devices in it. Various kinds of authentic language data, i.e., written texts, spoken conversations and corpus data were used to pin down the patterns of different referential forms (Pajusalu 1996, 1997a, 1997b, 1997c, 1999, 2000). Renate Pajusalu has also written the latest, up-to-date comprehensive overviews about the system of Estonian referential devices (Pajusalu 2009, 2017b).

In addition to Pajusalu’s works, the matter of text coherence and cohesion in upper secondary school students’ written essays is under discussion in Kersti Lepajõe’s (2011) dissertation, which adopts the methods of linguistic and functional text analysis. Lepajõe (2011: 43–50) analyses the array of different cohesive devices employed in upper secondary school students’ writing, including the use of lexico-grammatical textual devices for connecting referents, as well as discourse connectors for marking the relations between different parts of a text. In addition, Lepajõe (*Ibid.* 51–60) is interested in the relation of coherence to information structure, especially in terms of how it is reflected in the presentation of new vs. old information. Furthermore, Lepajõe (2004, 2008) has been interested in the topic of person deixis in student writing.

After these initial insights, studies started to move from a general point of view towards more specific studies about particular aspects of reference. Various starting points characterise the research tradition of Estonian referential devices. To start with, the traditional division between **anaphoric** (endophoric) and **deictic** (exophoric) reference also distinguishes two strands of research regarding Estonian. It seems that anaphoric reference (Pajusalu 1997c, 2000, 2005, 2006a; Kaiser and Hiietam 2003; Kaiser and Vihman 2010; Kirsipuu, Soodla, and Pajusalu 2012) has been studied more, but more recently, deictic devices, especially the functions of demonstrative pronouns, have also gained attention (Reile 2015, 2016, 2019; Pajusalu et al. 2018; Reile et al. 2019, 2020).

Estonian reference studies can also be divided into two branches based on the **types of NPs** that are investigated. There are studies investigating the usage contexts and frequencies of article-like determiners, such as definite *see* ‘this’ (Pajusalu 1997c; Hiietam 2003) and indefinite *üks* ‘one/a’ and *mingi* ‘some’ (Pajusalu 2000). Another group of studies focuses on the referential properties of pronouns. Demonstrative and personal pronouns have been analysed, for example, in standard Estonian (Pajusalu 1997a, 1997b, 2005) as well as in Võro (Pajusalu 1998, 2006; see also Reile et al. 2020). As there is an interesting variation of demonstrative systems among different varieties of Estonian, the aspect of dialectal variation has gained some attention in deictic (Pajusalu 2006a; Tammekänd 2015; Reile 2015, 2016; Reile et al. 2019) and anaphoric contexts (Tirkkonen 2007), or in both (Pajusalu 2015). The use of Estonian

pronouns (short vs. long forms and overt vs. zero reference) in the speech of Estonians living in America has also been studied (Kivik 2010).

Another vein of research has taken a **contrastive approach** and compares different linguistic devices that share some similar referential or formal properties. The comparisons are often about two different devices within Estonian; for example, there are studies that compare the indefinite determiners *üks* ‘one’ vs. *mingi* ‘some’ (Pajusalu 2000), the demonstratives *see* ‘this’ vs. *too* ‘that’ (Pajusalu 2006a; Reile 2019) or the short vs. long form of the third person pronoun (Kaiser 2010). In addition, Estonian referential devices are often compared with Finnish referential NPs. For example, Kaiser and Hiietam (2003) conducted a comparison of Estonian and Finnish third person anaphors in which they studied the effect of three factors (grammatical role, word order, and clausal organization) on referential choice. Kaiser (2010) also contrasts the usage contexts of Estonian short (*ta*) and long (*tema*) pronoun forms to Finnish pronouns in a small corpus of written texts. Another contrastive study is by Nordlund et al. (2013), in which article-like determiners in old literary Estonian and Finnish are analysed. The studies DETERMINERS and PRONOUNS in this thesis also contribute to this research tradition.

Referential practices in **child and caregiver language** have also gained some attention. For example, the expression of reference in child language and its resemblance to an adult system (Kirsipuu, Soodla, and Pajusalu 2012; Hint, Reile, and Pajusalu 2013), referential devices in Estonian child-directed speech (Vihman 2015), and determiners in child language (Vija and Pajusalu 2009; Hint, Reile, and Pajusalu 2013) have been of interest.

Estonian reference studies also exhibit **methodological diversity**. Qualitative studies (Pajusalu 1996, 1997a, 1997b, 1998, 2000), as well as corpus-based analyses have been carried out (Pajusalu 1997c; Kaiser and Hiietam 2003; Kaiser 2010). Also, a few psycholinguistic experiments have been conducted that directly test the effect of a couple of factors on certain referential expressions, for example, Kaiser and Vihman (2010) or Reile (2015, 2016) and Reile et al. (2020). There are also attempts to understand pronominal co-reference resolution in natural language processing, for example, by applying constraint grammar (Puolakainen 2015) or neural networks (Freienthal 2020). Still, some methodological advances have not been adapted to their full potential until now. For example, there are not yet any multifactorial corpus studies conducted that investigate the use of Estonian referential devices by applying up-to-date statistical techniques. The fourth study in this thesis, DEMONSTRATIVES, can be seen as a step towards filling this gap.

Many studies, although not directly focused on the study of reference, add valuable **nuances** to the overall knowledge about Estonian referential system. For example, Lindström and Vihman (2017) have analysed the marking of experiencer arguments and the optional omission of the oblique argument in ‘need’-constructions, which relates to the usage patterns of Estonian zero reference. Observations about the preferences for short vs. long forms of personal pronouns have been made in studies that concentrate on syntactic aspects. For example,

Lindström (2001) has discussed verb-initial clauses in narratives and demonstrated that if such clauses mark the subject of the clause with a pronoun, then it tends to be a long form (example 4). Pajusalu (2006b), on the other hand, has shown the preference for short pronoun in relative *wh*-constructions (example 5). In the next sections, these and other similar studies are also briefly introduced.

- (4)    *lähen*            *mina*            *tööle*            *ja*            *hakkab*            *sadama*  
          go.PRS.1SG    1SG.LONG.NOM    work.ALL    and    begin. PRS. 3SG    rain.INF  
          ‘I go to work and it starts raining’
- (5)    *mis*            *sa*                            *tast*                            *ootad*  
          what    2SG.SHORT.NOM    3SG.SHORT.ELA    wait.PRS.2SG  
          ‘why are you waiting for him’

The above-mentioned studies share a common starting point: they focus on describing the general usage patterns and referential properties of one particular device, or they compare the usage patterns of a few similar devices. This thesis adopts a different vantage point and thus adds yet uncovered aspects to the overall description of the Estonian system of referential devices. Instead of focusing on a single device and its functions in discourse, the thesis studies the behaviour of different referential devices in reference chains. Put differently, instead of putting a single referential device at the centre of attention, the thesis examines how different kinds of referents are referred to. The studies presented in this thesis observe which variables affect the choice of a referential device. The aim is to show that, first, different types of referential NPs may have similar referential potential in discourse (i.e., they can refer to similar entities), but their occurrence in discourse is affected by different variables. Second, similar types of referential devices might have different functions across languages. The thesis covers the whole range of possible devices for anaphoric reference: NPs with indefinite and definite determiners, demonstrative pronouns, personal pronoun forms, zero reference, and full NPs without a determiner.

### 3.2. Grammatical features that shape the Estonian referential system

This section gives an overview of the aspects of Estonian grammatical structure that are relevant to the use of referential NPs in particular.

**Case and agreement.** Referential acts are usually performed by NPs. Estonian is a morphologically rich language in which NPs can inflect through 14 case forms. Determiners and adjectives as NP attributes usually agree with the head noun in case (example 6). However, there are also instances of NP attributes which do not exhibit case agreement. Such attributes are nouns or pronouns that

function as genitive attributes (example 7) and adverbial attributes (example 8) (Pajusalu 2017a). In terms of reference, explaining the referential functions of NP attributes with different agreement patterns is more complicated. For an adequate explanation, it must be decided whether such NPs without direct agreement contain several referents or create a single, unified referent. For example, the NP *lapse raamat* ‘a child’s book’ involves two entities – a child (*laps*) and a book (*raamat*) –, which can each have a referential value of its own in the speech setting. However, it is in principle possible that the speaker, by uttering such a phrase, only refers to a book that has a property of belonging to a child, but the exact identity of the child is not important. As such, it is possible that in some contexts, non-agreeing NPs combine a referential as well as generic meaning. When a pronoun is used in such constructions instead of a full noun, the second interpretation is not available.

(6) NP case agreement

<i>ühe-le</i>	<i>suure-le</i>	<i>koera-le</i>
one-ALL	big-ALL	dog-ALL

‘to a big dog’

(7) genitive attribute

a) *lapse*                      *raamat*  
 child.GEN      book.NOM  
 ‘a child’s book’

b) *tema*                      *raamat*  
 he.GEN      book.NOM  
 ‘his book’

(8) adverbial attribute

a) *laul*                      *armastuse-st*  
 song.NOM      love-ELA  
 ‘a song about love’

b) *laul*                      *tema-st*  
 song.NOM      he-ELA  
 ‘a song about him’

Case clearly has the potential to affect referential choice in Estonian. For example, the choice between short and long pronoun forms (*ta* vs. *tema*) can depend on the case form of the NP (see section 3.3.3). Also, a particular case form can be determined by the construction in which the NP occurs (e.g., the elative *mis*-construction; see example 5 in section 3.1), and the use of a short pronoun form in elative case can pragmatically mark the whole construction (Pajusalu 2009: 125).

There is one type of referential phrase, though, which does not take case inflection at all, namely, the demonstrative adverbs *siin* ‘here’ and *seal* ‘there’. However, spatial demonstrative adverbs still have three different forms that reflect the functions of three spatial cases in Estonian: *siia* ‘to here’ and *sinna* ‘to there’ for GOAL referents, *siin* ‘here’ and *seal* ‘there’ for expressing LOCATION, and *siit* ‘from here’ and *sealt* ‘from there’ for SOURCE referents (Pajusalu 2009).

**Free word order.** Estonian has relatively free word order with varying patterns across different clause types (Erelt 2009). This means that referential devices have no fixed position in a clause and their placement depends on the purpose of the clause and the referential functions of an NP. The unmarked word order in basic clauses is SVO/SVX (Erelt 2009; Dryer 2013c; Lindström 2017: 547), but XVS and VS clauses are also common (Lindström 2005, 2017). The most important feature of Estonian word order is considered to be the V2 principle: the finite verb generally occupies the second position in a simple clause or main clause (Lindström 2017: 551). Recent studies also confirm that the V2 tendency is predominant in Estonian affirmative, declarative main clauses, especially in written register; however, more variation in verb placement occurs in spoken Estonian, in which V3 sentences are also common (Vihman and Walkden 2021).

The word order of a clause is determined by its information structure. The topic, i.e., old information, is typically presented at the beginning of a clause, and the focus, i.e., new information, is given towards the end of a clause (Lindström 2017: 549; Tael 1988). Previous research has shown the effect of word order on the choice of specific referential expressions in Estonian. For example, Kaiser and Vihman (2010) have demonstrated that the use of anaphoric demonstrative pronouns is sensitive to the word order of a preceding clause, while personal pronouns do not exhibit such a preference.

**Syntactic relations.** The choice of a referential expression is closely related to its syntactic role in the clause. A crosslinguistic tendency seems to be that grammatical subjects exhibit higher salience than other roles and thus anaphoric pronouns prefer the syntactic subjects of a preceding clause as their antecedents, while longer referential forms are used for non-subject arguments (Gordon, Grosz, and Gilliom 1993; Crawley, Stevenson, and Kleinman 1990; Du Bois 2003). In Estonian, this means that personal pronouns and zero reference should be expected to evoke subjects (see also Metslang 2013), while more elaborate forms more often mark objects and other arguments in clauses. However, there are some clause types in Estonian in which the grammatical subject is not the most agentive argument in the clause (Erelt et al. 1993: 14) and instead, other arguments express some subject properties (Metslang 2013). This holds true for marked basic clauses (i.e., existential, possessive, source-marking resultative, and experiential clauses) in which the main topic takes the position of an adverbial, direct object or an oblique (Erelt and Metslang 2006). This feature of Estonian has important implications on reference resolution, as became especially clear when applying Centering theory to Estonian data (ZERO).

**“Missing” features.** In describing the referential paradigm of Estonian, features that do not belong to the Estonian grammatical system must also be kept in mind. First, Estonian has no grammatical gender. In the context of the present thesis, PRONOUNS contrasts the Estonian referential system with the Russian system, in which grammatical gender occurs.

Second, there are no grammatical articles, neither definite nor indefinite, in Estonian (Erelt 2017b: 76; Dryer 2013a, 2013b; see also Haspelmath 1998). However, the Estonian system includes several determiners that function very similarly to grammatical articles: for example, Estonian detNP *see mees* can be translated into English as ‘this man’ or ‘the man’. Such use of determiners can be described in terms of grammaticalization (Heine and Kuteva 2006), in that certain determiners have been observed to gradually develop into grammatical articles (e.g., Diessel 1999; Givón 1981; Greenberg 1978; C. Lyons 1999; Weiss 2004). The article-like properties of determiners in Estonian have been discussed for example in Pajusalu (1997c, 2000), but the presence of “real” grammatical articles in Estonian has not been proposed.

These features are not remarkable on their own, but it is important to take them into account when we compare the Estonian referential paradigm to other, typologically different languages. Therefore, when coding and analysing these languages in parallel, adjustments must be made with regard to the coding and analysing processes.

The following sections of this chapter give a description of more specific properties of individual Estonian referential devices.

### 3.3. Referential devices in Estonian

This section is intended as a synopsis of the main types of referential NPs that are commonly used for anaphoric reference in Estonian. In Table 1, all of the described Estonian referential devices and corresponding examples are presented. It must be pointed out, however, that this thesis does not intend to give an exhaustive list of all possible referential devices in Estonian. Other kinds of more complex referential phrases are also possible, for example, relative clauses (Pajusalu et al. 2018), initial and final detachments (e.g., Amon 2015) or postponed demonstratives (Yurayong 2020). In this thesis, the focus is only on NPs.



**Table 1.** Referential NP types in Estonian and corresponding examples (\* in certain Estonian regions only)

	Devices	Forms	Example
Most reduced	Zero reference	ø	<i>Mees sõi saia, siis ø jõi kohvi.</i> Man.NOM eat.PST.3SG bread.PRT, then ø drink.PST.3SG coffee.PRT.  ‘A man ate bread, then ø drank coffee.’
	persPron	short <i>ta</i> ‘s/he’	<i>Ta sõi saia, siis jõi ta kohvi.</i> 3SG.SHORT eat.PST.3SG bread.PRT, then drink.PST.3SG 3SG.SHORT coffee.PRT  ‘He ate bread, then he drank coffee.’
		long <i>tema</i> ‘s/he’	<i>Ta<sub>1</sub> sõi saia, aga tema<sub>2</sub> jõi hoopis kohvi.</i> 3SG.SHORT eat.PST.3SG bread.PRT, but 3SG.LONG drink.PST.3SG instead coffee.PRT  ‘He <sub>1</sub> ate bread, but he <sub>2</sub> drank coffee instead.’
	dem	pronoun <i>see</i> ‘this/that’ * <i>too</i> ‘that’	<i>Laual oli sai. Mees sõi seda natuke.</i> Table.ADE be.PST.3SG bread.NOM Man.NOM eat.PST.3SG it.PRT a bit  ‘There was bread on the table. A man ate it a bit.’
		adverb <i>siin</i> ‘here’, <i>seal</i> ‘there’ etc.	<i>Laual oli klaas. Seal oli ka sai. Mees sõi seda.</i> Table.ADE be.PST.3SG glass.NOM. There be.PST.3SG also bread.NOM. Man.NOM eat.PST.3SG it.PRT.  A glass was on the table. There was also a bread. A man ate it.’
	detNP	definite det. <i>see</i> ‘this’, <i>too</i> ‘that’, <i>oma</i> ‘one’s’ etc.	<i>See mees sõi seda saia.</i> DEM.NOM man.NOM eat.PST.3SG DEM.PRT bread.PRT.  ‘The man ate this bread.’
indefinite det. <i>üks</i> ‘a/one’, <i>mingi</i> ‘some’ etc.		<i>Üks noor mees sõi mingit saia.</i> One.NOM young.NOM man.NOM eat.PST.3SG some.PRT bread.PRT.  ‘A young man ate some kind of bread.’	
Least reduced	fullNP w/o a determiner	N adj + N etc.	<i>Mees sõi maitsvat saia.</i> Man.NOM eat.PST.3SG tasty.PRT bread.PRT.  ‘(A) man ate tasty bread.’

### 3.3.1. Estonian article-like determiners

Estonian is commonly regarded as a language without grammatical articles (Dryer 2013a, 2013b; Erelt 2009; Norris 2018). However, several pronouns can be used as determiners for expressing certain referential properties of an NP (Pajusalu 2017a, 2017b). In Estonian, determiners commonly precede the head noun (*see nõid*, as in example 10), although in rare occasions, other positions are also possible (e.g., postponed demonstratives, as in *rong see sõitis* ‘train the drove’; see Yurayong (2020) for a detailed overview of this phenomenon). Both definite and indefinite determiners are attested in natural linguistic discourse. For example, when a speaker introduces a referent for the first time in a discourse, she might indicate this newness by an indefinite pronoun *üks* ‘one’ (9), but when the same referent is mentioned again in the next clause, the referent’s knownness can be indicated by a definite determiner *see* ‘this’ (10).

- (9) *Elas*                      *kord*              *üks*              *nõid*  
live.PST.3SG      once              one/a              witch.SG.NOM  
‘Once upon a time there lived **a witch**.’

- (10) *See*                      *nõid*                      *oli*                      *väga*              *õel*.  
This/the              witch.SG.NOM              be.PST.3SG              very              mean.NOM  
‘**This/The witch** was very mean.’

No finite list of all possible determiners in Estonian has been proposed, but pronouns and pro-adverbs such as *see* ‘this’, *too* ‘that’ *selline* ‘that kind of’, *siin* ‘here’, *seal* ‘there’, etc. are more commonly listed as definite determiners, and the pronouns *üks* ‘one’, *mingi* ‘some’, *keegi* ‘someone’, *miski* ‘something’, and *kuskil* ‘somewhere’ as indefinite determiners (Pajusalu 2017b). Previous studies regarding Estonian determiners have focused on the most common determiners, namely the demonstrative pronoun *see* ‘this’ as definite determiner and the pronouns *üks* ‘one’ and *mingi* ‘some’ as indefinite determiners (Pajusalu 1997c, 2000, 2005, 2009). These studies have described the categories, both grammatical and pragmatic, that are relevant to the referential choice.

While some determiners serve a specific meaning, such as the pronoun *oma* ‘one’s’ for expressing possession, or are only common to certain contexts, such as *mingi* ‘some’ in an informal context (Pajusalu 2000), others occur more freely and seem to have fewer restrictions. This is especially true when the demonstrative pronoun *see* ‘this’ is used in a determiner position or when the indefinite pronoun/numeral *üks* ‘one’ behaves as an indefinite determiner. As these tendencies are attested in a wide range of languages, they are often considered to be universals across languages (Greenberg 1978; Diessel 1999; Heine and Kuteva 2006). Following these studies, Nordlund et al. (2013) investigated the use of article-like determiners in old literary Estonian and Finnish and found that article-like functions, which were not only translational calques from

German and Swedish, were already present in 16<sup>th</sup> century texts in both Estonian and Finnish.

Importantly, in addition to explanations pertaining to referential function, other views on the uses of demonstrative pronouns in a (seemingly) determiner position have also been presented. For example, Sahkai (2003) has offered an alternative and explained the uses of adnominal demonstratives as demonstrative doubling in spoken Estonian. A different, interactional view is presented in Keevallik (2010), where demonstratives are seen as placeholders, at least in spoken discourse. Therefore, it is plausible that (adnominal) demonstratives have various functions, of which the article-like function is just one possible interpretation among others.

However, the use and function of article-like determiners in Estonian has not yet gained more systematic attention, and thus, many questions remain unanswered. Firstly, for example, the relative frequencies of different determiners in Estonian discourse as compared to frequencies of other referential devices is not documented. In addition, the factors that affect the choice of a particular device are not well known. Also, it is not clear whether these frequencies are comparable to usage patterns in the other article-less contact languages of Estonian. It is hypothesized that such a comparison might be informative regarding the level of grammaticalization of article-like determiners. Study PRONOUNS in this thesis emanates directly from these problems.

Secondly, the subtle differences between similar determiners have not been established. In this thesis, two constructions that include definite determiners are compared – one with demonstrative pronouns *see* ‘this’ and *too* ‘that’ and the other with demonstrative adverbs *siin* ‘here’ and *seal* ‘there’ (DETERMINERS).

### 3.3.2. Estonian demonstratives as independent devices

Demonstratives, being widespread referential devices across languages, are frequently used for anaphoric referral in Estonian as well. The system of Estonian demonstrative pronouns is remarkable since the number of demonstrative pronouns depends on the regional variety spoken. Three different demonstrative systems have been attested (Pajusalu 1996, 2006a; see also Reile 2015, 2016). In Northern varieties and Standard Estonian, only one, distance-neutral demonstrative pronoun, *see* ‘this’, is used. While the distal demonstrative *too* ‘that’ can also occur in the Standard variety, it is usually restricted to temporal expressions (Pajusalu 2006a; Reile et al. 2019). A two-term demonstrative system is observed in Southern varieties of Estonian, where proximal *see* ‘this’ and distal *too* ‘that’ are commonly used. Moreover, in some dialects, more than two demonstrative stems are used, for example *sie*, *tämä* and *tä* in North-Eastern and Coastal dialects (Tirkkonen 2007) and *sjo*, *taa* and *tuu* in Võro (Pajusalu 1998; Tammekänd 2015). If more than one demonstrative is used, then *see* is considered proximal and *too* distal (Pajusalu 2006a; Reile 2019). Regardless of the variety, demonstrative pronouns can be used either pronominally (e.g., *see*

*haiseb* ‘it/this stinks’) or adnominally (e.g., *see kala haiseb* ‘this fish stinks’, see also section 3.3.1). This section is devoted to the pronominal uses.

As in other languages, anaphoric as well as deictic reference can be conveyed with demonstrative pronouns in Estonian. Regarding Estonian demonstrative pronouns, both anaphoric (e.g., Hint et al. 2013; Kaiser and Vihman 2010; Pajusalu 1997a, 2005, 2006a) and deictic (e.g., Pajusalu 2006; Reile 2015, 2016, 2019; Reile et al. 2019) uses of demonstratives have been studied. In this thesis, the focus is on the anaphoric functions of Estonian demonstratives. However, this thesis does not focus on the properties of a particular referential device alone, but, differently from the previous studies, it looks how different referential devices work together in discourse, and which factors influence referential choice.

The anaphoric demonstrative pronoun *see* ‘this’ typically refers to inanimate and abstract entities in discourse (Pajusalu 2017b: 575). However, it is not uncommon for the demonstrative *see* to be used for referring to animate entities as well. Three main explanations have been given for such uses. First, the demonstrative *see* is characteristic of contexts in which the animate referent’s cognitive status is not (yet) the highest, i.e., in focus (Kaiser and Hiitam 2003; Pajusalu 2009). Second, demonstratives are useful in contexts in which two different animate entities, both salient, are mentioned within one clause. In such cases, the personal pronoun probably refers to first entity mentioned (usually the subject), whereas the demonstrative is restricted to the second entity (usually the object) in a clause (Erelt et al. 1993: 209). Third, the speaker’s negative attitude towards the referent might trigger the use of a demonstrative pronoun for referring to persons (*Ibid.*).

The asymmetry between *see* and *too* is an explanation for the fact that studies observing the functions of Estonian *too* are rare. As has been previously noted, *too* is more restricted in its possible uses. These uses are mostly related to temporal or more distance-based meanings (Pajusalu 2006a; Reile et al. 2019; Taremaa et al. 2021).

Besides demonstrative pronouns, demonstrative adverbs also function as referential devices. In Estonian, there are six demonstrative adverbs: three proximal and three distal. Proximal demonstrative adverbs are *siia* ‘to here’, *siin* ‘here’ and *siit* ‘from here’, distal demonstrative adverbs are *sinna* ‘to there’, *seal* ‘there’ and *sealt* ‘from there’. Demonstrative adverbs are applicable in contexts in which spatial meanings are expressed: thus the referential functions of demonstrative adverbs are restricted in Estonian since they usually refer to places or behave like definite determiners. For example, in (11), the demonstrative adverb *seal* ‘there’ clearly has a referential function, as its purpose is to re-evoke the university as a specific (place-related) entity.<sup>9</sup>

---

<sup>9</sup> Locations can be considered as a different kind of concepts than referents. However, conceptualizing locations as referents is a common practice in languages (Kibrik 2011: 7).

- (11) *Peeter*            *töötab*            *ülikoolis.*  
 Peter.NOM        work.PRS.3SG        university.INE  
*Seal*            *töötab*            *ka*        *Mari.*  
 There            work.PRS.3SG        also        Mary.NOM  
 ‘Peter works at the university. Mary also works **there.**’

### 3.3.3. Estonian third person personal pronouns

In the paradigm of Estonian personal pronouns, each grammatical person (i.e., first, second, and third) can be expressed either with a short or long form.<sup>10</sup> The third person forms are short *ta* (example 12) or long *tema* ‘s/he’ (example 13) in singular, and short *nad* or long *nemad* ‘they’ in plural. The unmarked, most neutral and phonetically unstressed anaphoric form for referring to most salient discourse entities is the 3SG short form *ta* ‘s/he’ (Pajusalu 2009).

- (12) *Peeter*            *töötab*            *ülikoolis,*  
 Peter.NOM        work.PRS.3SG        university.INE  
*ta*                    *on*                    *professor.*  
 3SG.SHORT.NOM    be.PRS.3SG        professor.NOM  
 ‘Peter works at the university, **he** is a professor.’
- (13) *Mari*            *töötab*            *ka*        *ülikoolis,*  
 Mary.NOM        work.PRS.3SG        also        university.INE  
*tema*                *on*                    *lektor.*  
 3SG.LONG.NOM    be.PRS.3SG        lecturer.NOM  
 ‘Mary also works at the university, **she** is a lecturer.’

It is suggested that the distinction between short and long forms is not only rooted in grammar, but it is often a pragmatic choice as well (Pool 1999; Erelt et al. 1993; Pajusalu 2009). For example, according to current knowledge, the most important factor that leads towards the use of a long form is contrast (Pajusalu 1997b). Kaiser and Hiietam (2003) have also shown that while the Estonian short form *ta* usually refers to subjects of preceding clauses, the long form *tema* is rather sensitive to contrast with other referents. Kaiser (2010) adds that the contrast-based account explains the distinction between Estonian short and long pronoun forms more clearly than the salience-based account.

However, contrast is not the only factor that lies behind the short and long forms. For example, in certain syntactic constructions, particular forms are preferred. In spoken narratives, verb-initial clauses may prefer long pronoun forms (Lindström 2001; see also example 4 in section 3.1). In addition, the case

<sup>10</sup> The terms ‘weak’ and ‘strong’ pronouns have also been used in literature, e.g., Kaiser (2010).

form can affect the choice between short and long forms. In genitive, inner-location cases and also comitative, the long form is preferred (e.g., GEN *tema*, INE *tema-s*, COM *tema-ga*). The short form, on the other hand, is more frequent among nominative and outer-place case forms (e.g., ADE *ta-l*) (Pajusalu 1997a, 2017b; see also Pool 1999 for first and second person forms). Notably, the semantic cases translative, terminative, essive and abessive, and the partitive case lack the short form overall.

A referent's inherent properties are also relevant to pronoun choice. Personal pronouns, both long and short, usually refer to animate entities in discourse (Pajusalu 2009, 2017b). This is especially clear in narrative contexts, in which animate entities who perform a central activity in a narrative often occupy a prominent role in longer reference chains (Pajusalu 2017b). However, inanimate entities may also be referred to using a personal pronoun in certain contexts (Erelt et al. 1993: 207–208; Pajusalu 2017b). This can be attested, for example, when an inanimate entity is brought into the centre of attention and is referred to in more than two consecutive sentences. Thus, the speaker's decision to use a personal pronoun for an inanimate referent is also a pragmatic, discourse-related move (Pajusalu 2009).

In sum, the choice of a pronoun form has been explained based on individual factors, such as contrast, case form, animacy, salience, etc. This thesis takes a multifactorial perspective instead and examines the impact of several factors working together simultaneously. Also, the relative position of personal pronouns within the overall referential system of Estonian, as compared to other referential devices, is considered. In addition, a specific comparison is made with another referential device, namely zero reference (study ZERO), which shares many similar properties with the short forms of third-person pronoun but had not previously gained sufficient attention in Estonian referential descriptions.

### 3.3.4. Zero reference in Estonian

In this thesis, the term 'zero reference' is employed to mark the lack of an overt referential form, i.e., the absence of overt argument of the verb in contexts in which it is expected. In example (14), adopted from the first study ZERO, the referent is evoked using the overt third person pronoun *ta*, but in a similar passage from another speaker (example 15, adopted from ZERO), zero reference is used instead. Examples (14) and (15) show that the overt form *ta* and zero reference exhibit important similarities in their referential properties (i.e., highly salient animate referent, no other animate entities in discourse, etc.) and can thus be analysed as near-synonymous constructional variants.

- (14) *seal oli mees farmer või keegi*  
 there be.PST.3SG man.NOM farmer.NOM or somebody.NOM  
*jaa ta korjas puul ta*  
 and 3SG.SHORT.NOM pick.PST.3SG tree.ADE 3SG.SHORT.NOM  
*korjas pirne*  
 pick.PST.3SG pear.PL.PRT  
 ‘there was a man, a farmer or somebody. and **he** was picking, up on a tree, **he** was picking pears’
- (15) *oli üks mees kes korjas pirne*  
 be.PST.3SG one/a man.NOM who pick.PST.3SG pear.PL.PRT  
*jaa ø korjas pirne rahulikult*  
 and ø pick. PST.3SG pear.PL.PRT calmly  
 ‘there was a man who was picking pears. and **ø** was picking pears calmly’

Therefore, it is possible to refer to highly salient discourse entities without explicitly mentioning the referent with an overt NP. In such patterns, some researchers prefer to avoid talking about the zero pronoun in an absolute sense, since the verb inflections usually mark the person in Estonian (e.g., PRS 1SG *kirjuta-n* ‘I write’, PRS 2SG *kirjuta-d* ‘you write’, PRS 3SG *kirjuta-b* ‘she writes’), so it is possible to infer the salient referent based on the verb form itself (Siewierska 2004; Kibrik 2011). This may be one reason why earlier treatments of the Estonian referential system have not included zero reference as a separate device (Erelt et al. 1993; Pajusalu 2009), but have explained similar clauses under the wider term ellipsis, which stands for the phonetically zero marking of whichever clausal component (Erelt et al. 1993).

Previous research regarding Estonian zero reference has mostly been concerned with first and second person subjects. For example, Duvallon and Chalvin (2004) have analysed first and second person reference, and Lindström et al. (2008) have studied the use of first person for dialect data. There are also studies in which all three persons are studied in comparison (Kivik 2010; Vihman 2015; Lindström and Vihman 2017). These studies agree that zero reference is most common in second person (Kivik 2010; Lindström and Vihman 2017; Duvallon and Chalvin 2004), while both first and third person can occur with either zero or overt reference (Kivik 2010; Lindström and Vihman 2017). The existing research regarding Estonian zero reference has also concentrated on the syntactic properties of omitted arguments, for example in oblique experiencer constructions (Lindström and Vihman 2017). Also, the expression (and omission) of arguments in Estonian child-directed speech has been studied (Vihman 2015). In these studies, researchers have preferred the notion ‘argument omission’ instead of zero reference.

In sum, it has been concluded that Estonian third person zero reference is used less frequently than first and second person zero reference, and it is only licensed by strong contextual cues (Vihman 2015). For example, Estonian zero

reference is only common with highly ranked subjects (Lindström et al. 2008; Metslang 2013) and is considered acceptable in subsequent sentences sharing a similar context and common subject (Erelt et al. 1993: 225; Lindström 2005). Also, Estonian third person zero reference has mostly been attested in (spoken) narratives (Lindström 2005; Keevallik 2003).

The current thesis (especially studies ZERO and PRONOUNS) is the first attempt to precisely compare the overt third person form *ta* and the zero reference in Estonian narrative data. In this thesis, a central question is which factors influence referential choice in Estonian? More specifically, is there a clear difference between contexts in which an overt third person form is used, compared to the contexts in which zero reference occurs? Also, this thesis places zero reference into the overall Estonian referential paradigm and compares the usage patterns of Estonian zero reference to those of Finnish and Russian (PRONOUNS).

### **3.3.5. The relationship between different types of NPs**

The types of Estonian referential devices that are the focus of the current thesis are arranged in Table 2, which takes into account the distinctive features of referential devices discussed in Chapter 2 (section 2.2), i.e., full vs. reduced referential devices and definite vs. indefinite referential devices. The horizontal axis separates definite NPs from indefinite ones, and the vertical axis describes the reduction of a referential device. The category of definiteness is captured by the formal properties in Table 2. However, while the importance of definiteness in Estonian referential system is acknowledged, it should be noted that such a formal division has certain limits, especially when applying it to natural discourse data. When we turn to the following chapters of this thesis, it becomes clear that not all of these devices appeared in the data collected for the thesis (e.g., *tema*, *keegi*, and also independent demonstrative adverbs). It is unfortunately out of the scope of this study to explore all the devices with the diligence that they deserve. Still, their place in the referential system is recognised.



**Table 2.** Types of Estonian referential devices by definiteness and reduction

Indefinite NPs		Definite NPs
Full	NPs w/o determiner e.g., <i>maja</i> ‘house’ <i>suur punane maja</i> ‘big red house’	
	NPs with an indefinite determiner e.g., <i>üks raamat</i> ‘a book’	Proper nouns e.g., <i>Miisu</i>
Reduced	Indefinite pronouns e.g., <i>keegi</i> ‘somebody’	NPs with a definite determiner e.g., <i>see raamat</i> ‘the/this book’ <i>see raamatus</i> ‘in this book [there]’ <i>minu arst</i> ‘my doctor’
		Demonstratives pronouns, e.g., <i>see</i> ‘this’ adverbs, e.g., <i>siin</i> ‘here’
		Personal pronouns <i>ta</i> ‘s/he’ short form <i>tema</i> ‘s/he’ long form
		Zero reference

In sum, there are many types of NPs in languages, and as exemplified based on the Estonian language in this thesis, these NPs can serve a referential function in discourse. Importantly, this is not an exhaustive list of referential devices. More complex nominal constructions, for example longer modified NPs (Ariel 1990) and relative clauses (Pajusalu et al. 2018; Hoek et al. 2020), or even certain grammatical functions, such as case or the passive construction (Givón 2005: 135) or agreement (Kibrik 2019), can also be utilised for coding referents. As complex NPs are rare in natural discourse (Kibrik 2011: 38), they are often neglected in works that aim to present a concise system of referential devices in a language.<sup>11</sup> The exact factors on the basis of which the speaker chooses (and the addressee interprets) different referential forms is of central interest in this thesis.

<sup>11</sup> This thesis also disregards complex referential NPs from the analysis and discussion. Refer to Pajusalu et al. (2018) for an analysis of relative clauses as referential devices in Estonian, Finnish, and Russian.

## 4. METHODOLOGY

This chapter describes the data collection and analysis methods used in the four studies presented in this thesis. First, the data collection techniques are explained in section 4.1. An elaborated list with descriptions regarding all the explanatory variables included in the studies is given in section 4.2. Section 4.3 presents an overview of the applied statistical analysis methods.

### 4.1. Data collection methods

In this thesis, data from spoken as well as written language are examined. For collecting spoken language data, quasi-experimental methods were applied (ZERO, DETERMINERS, and PRONOUNS). Written data were gathered from a corpus (DEMONSTRATIVES). As became evident while conducting the studies reported in this thesis, corpus data is also necessary for analysing certain aspects of reference. In this light, the variety of data used has also been one of the methodological considerations in the study of reference, which are further discussed in Chapter 6. While the focus in this thesis is very much on (informal) spoken discourse, its scope also extends to the written mode.

In Table 3, an overview of all different types of data used in the thesis is presented. In the sections below, these datasets are described in more detail.

**Table 3.** Overview of the data used in the thesis

Study	Data	Lan- guages	No. of Participants / No. narratives	No. of analysed NPs
ZERO	Spoken narrative (Pear Film)	Estonian	16/16	517
DETERMINERS	Spoken narrative (picture sequences)	Estonian	20/59	676
		Finnish	20/54	680
PRONOUNS	Spoken narrative (picture sequences)	Estonian	20/59	540
		Finnish	20/54	464
		Russian	20/57	530
DEMONSTRATIVES	Written corpus sentences	Estonian	–	2400

#### 4.1.1. Spoken narrative data

Using narrative data for different purposes in linguistics as well as in other disciplines has a long history (e.g., Labov 1972; Labov and Waletzky 1967; Stein and Glenn 1979). One well-known and oft-cited project is “The Pear Stories” project<sup>1</sup> from 1980 (Chafe 1980), which was launched to find out how simple stories can vary across languages. Another highly influential project was based on “Frog where are you” narratives, which focused on child language (Berman and Slobin 1994). As for current-day linguistics, narratives as a means for studying language has not lost their importance, as myriad studies can attest (e.g., Dumont 2016; Helasvuo 1993; Hickmann 2004; Koster, Hoeks, and Hendriks 2011; Mäkinen et al. 2014; Reilly et al. 2004; Vliet 2009). Simple narrative data is also deemed best for the purposes of this thesis. In the following section, the reasons underlying this choice are explained.

Referential features are present in every form and medium of language. However, studies about reference very often find narrative data useful, since in coherent narratives, reference chains tend to be longer than just two or three mentions, thus making it possible to observe the alternation of different devices throughout the chain (e.g., Chafe 1980; Himmelmann 1996; Kibrik 1996). Narratives are characterized as a basic and pervasive discourse form in language and thus, the referential patterns emerging from narratives are considered to be instances of prototypical reference (Kibrik 1996: 259; Reilly et al. 2004: 230). Furthermore, narrative structure is closely linked to language, culture, and context, and thus, cross-linguistic variation and different typologies can be revealed by studying narratives (Berman and Slobin 1994; Hickmann 2004; Mäkinen et al. 2014). The diverse nature of narratives makes them good data for studying various aspects of linguistic structure (Chafe 1980; Berman and Slobin 1994), but in the context of this work, it is crucial that narratives are especially well suited for studying anaphoric referential devices.<sup>12</sup> For example, narrative data has already yielded good results in studying referential devices in Estonian child language (Kirsipuu et al. 2012).

The quasi-experimental approach to collecting narratives is preferred in this study since it enables one to collect referential devices produced in similar conditions and in a similar manner. In this way, we can ensure that the observable factors are comparable across discourses and across participants. Two settings for collecting the narratives are used. The aspects discussed above relating to narratives as a data source all hold for both of the settings used here. While some methodological differences emerge when comparing the study ZERO to the studies DETERMINERS and PRONOUNS, the overall nature of the final data is not greatly affected by these differences.

---

<sup>12</sup> It has been claimed that in fictional narratives, deictic devices, especially those related to time and space, are used in ways that are inconsistent with spoken discourse (Dancygier 2019). As this thesis focuses on anaphoric reference, no fundamental difference between patterns found in narratives and in spoken discourse is expected.

## Setting 1: Pear Stories

In the first study, ZERO, the Pear Film (Chafe 1980)<sup>13</sup> was used to collect spoken narratives from 16 female participants. University of Tartu mailing lists and social media (Facebook) were used to recruit participants. All experimental sessions were carried out in Tartu in different seminar rooms at one university building. All participants who volunteered were native speakers of Estonian. The original set-up described in Chafe (1980) was adjusted so that in the present experiment, a listener who was not informed about the content or the plot of the film was included in the setting. In the experiment, one participant (the speaker) was asked to watch the Pear Film. The other participant (the listener) was asked to wait behind the door. After watching the film, the listener was invited into the room and the speaker was requested to retell the Pear Film to the listener. The participants were invited to participate in pairs, so that the speaker and the listener were acquainted with each other. Thus, altogether 32 persons participated in the study, but only 16 of them were active participants, i.e., those who produced crucial linguistic data for the study. In that way, we eliminated situations in which the speaker would produce less natural and more formal narratives due to the possible intimidation of an unfamiliar person as the listener. The narratives were audio recorded and transcribed based on the recordings. The coding procedure was carried out by the author. The exact coding process is described in detail in ZERO.

The 16 collected narratives were coded for the referential expression (overt pronoun or zero reference) and tagged for the explanatory variables. This procedure yielded 517 third person reduced NPs: 238 overt forms and 279 occurrences of zero reference (see ZERO for more details).

## Setting 2: Picture-sequence narratives

For the studies DETERMINERS and PRONOUNS, a different narrative elicitation technique was used than in ZERO. Three simple picture-books with six pictures in each were originally created for the study to elicit short spoken narratives from native speakers of three different languages: Estonian, Finnish, and Russian. 20 participants from each language, 60 participants altogether, volunteered for participation in the study. Different social media platforms and mailing lists were used to recruit participants. Experimental sessions with Estonian speakers (13 females, 7 males) took place in Tartu and Tallinn, Estonia. Russian speakers' (18 females, 2 males) narratives were collected in Tallinn, Estonia. Sessions with Finnish speakers were conducted in Tampere and Kuopio, Finland, and in Tartu, Estonia.

---

<sup>13</sup> The film was shown from YouTube (<https://www.youtube.com/watch?v=bRNSTxTpG7U>). Permission to use the film was obtained from Prof Wallace Chafe in an e-mail correspondence.

The participants attended the elicitation session alone. They were shown the three books in a random order. After seeing each book, the participant was requested to tell a short story about the events in the book to someone who was not present at the time. Thus, each participant produced three short narratives, i.e., the initial number of narratives from each language is 60. The narratives were audio recorded, transcribed, and then double coded by two researchers (the authors of DETERMINERS and PRONOUNS). The more exact details regarding the data collecting and coding procedure of the picture-sequence based elicitation are thoroughly described in the corresponding studies.

### **Pear Stories vs. picture-sequence narratives**

The question may arise whether it was necessary to switch the narrative-elicitation method in the process of developing this thesis. The reasons why the Pear Film was not adopted for collecting the contrastive data from three different languages are as follows:

i) For collecting comparable data from several languages, it is preferable for the data-collection method to be as simple and straightforward as possible. For example, it might be difficult to show a motion picture in certain places and conditions (e.g., poor internet connection, lack of necessary technical devices, no access to a quiet room equipped with office desk and chair, etc.). These conditions might seem irrelevant in the context of modern Western society. However, if the goal is to promote a method that may be useful for studying *every* language, then these considerations must be taken into account. The picture-sequence based narrative elicitation method used in this thesis was developed in accordance with the aim to implement it in future studies about lesser-spoken Finno-Ugric languages and smaller Estonian dialects, for example. The areas where they are spoken may have variable access to technology.

ii) The Pear Film lasts almost six minutes, and therefore the narratives based on that are also quite long (all narratives were longer than 20 utterances). In such relatively lengthy narratives, the same referents perform several successive actions throughout the film. However, the referents themselves remain the same and consequently, longer reference chains emerge in Pear Stories narratives. However, for the purpose of the present thesis, it was worthwhile to collect more references to different referents who perform similar actions or appear in similar settings. For that reason, three picture sequences, each yielding a relatively shorter narrative than one Pear Story, were constructed. That made it possible to observe more referential acts uttered in similar contexts but for somewhat different referents.

iii) Referential chains elicited with the Pear Film were longer and contained more referential NPs, and they were produced in a more natural setting. Narratives elicited with the picture sequences can be characterised as being slightly less natural, and as they were shorter, reference chains were also shorter in picture-sequence narratives. The naturalness factor boils down to the well-

known issue of the presence of the researcher. It was already noted in the original Pear Stories experiment in the 1980s that the participant knows that the researcher has seen the film and knows the plot. This, in turn, might affect the way that narratives are told and structured (Du Bois 1980b: 6). In the Pear Stories setting used for this study, participants came to the experiment in pairs, so the person who had to tell the story now definitely knew that the listener had not seen the movie. The picture-sequence-based narratives, on the other hand, were told in a researcher-and participant-only environment, although an imaginary listener was brought into the discourse via the instruction (“*please tell the story so that someone who is not present at the moment can understand the plot without seeing the pictures*”).

#### 4.1.2. Written corpus data

In DEMONSTRATIVES, corpus data is used to explain the differences between demonstrative pronouns and demonstrative adverbs in the determiner position. As previous research has indicated (Pajusalu 1997a, 1997b, 2009, see also DETERMINERS), the demonstrative pronoun *see* ‘this’ as a determiner is a common referential device in Estonian. In contrast, the use of the demonstrative adverbs *siin* ‘here’ and *seal* ‘there’ as determiners is infrequent in spoken narrative data (DETERMINERS). However, relative infrequency in one type of data does not necessarily indicate the overall scarcity of this linguistic device in a language in general. Therefore, it is important to find different kinds of data that can reveal more precise behaviour of a particular device. Thus, in this study, corpus data was collected in order to have comparable instances of the two alternative constructions (demPronNP vs. demAdvNP).

For DEMONSTRATIVES, 2400 clauses were taken from the Estonian National Corpus 2017 via Sketch Engine. 1200 clauses represented demPronNP constructions (e.g., *selles majas* ‘in this house’) and 1200 demAdvNP constructions (e.g., *siin majas* ‘in this house [here]’). To obtain a balanced dataset, both constructions were divided into 12 sub-constructions according to the demonstrative type (proximal vs. distal) and grammatical case (illative, allative, inessive, adessive, elative, ablative). Exactly 100 clauses representing each sub-construction were included in the dataset. Refer to DETERMINERS for more detailed description of the corpus data.

Each clause was then manually annotated for several variables, three of which are used in the paper: semantic class of the noun, concreteness of the noun, and semantic type of the verb. All the variables included in the data analysis of this thesis are more elaborately described in the next section.

## 4.2. Description of explanatory variables

The multivariate nature of linguistic data is not restricted to studies on referential expressions only, of course. Many kinds of linguistic choices and patterns have been shown to be sensitive to several factors. The phenomenon of two or more alternative constructions for encoding similar meaning in a language is rather widely treated in linguistic research (e.g., Baayen et al. 2013; Bresnan 2007; Gries 2001, 2003; Klavan 2012; Szmrecsanyi and Hinrichs 2008), and the results indicate that important (language-specific) factors exist that distinguish these forms. For studying constructional variation, such as *demPronNP* vs. *demAdvNP* or zero vs. overt reference in Estonian, a multifactorial approach shows good results in this thesis, also.

A multifactorial approach to referential choice is adopted in this thesis and thus, several variables are annotated for each referential device. Depending on the study, not all variables are included, but they have all proven relevant in explaining certain aspects of referential choice in discourse. The relevant variables of each individual study are listed and described more thoroughly in the corresponding original article. This section presents a compendious list of all 13 linguistic variables that are investigated in this thesis.

The choice of the variables for each study was made based on the existing literature regarding referential choice in general and Estonian referential devices more specifically. The amount of all possible variables that can have an effect on the referential choice is, of course, vast. For example, in Kibrik et al. (2016), more than 20 candidate factors for the referential choice are listed. The factors also fall into different categories, such as grammatical factors related directly to the structure of language, factors related to the referent, factors pertaining to the discourse, etc. (Kibrik et al. 2016; Fukumura and van Gompel 2015; Järvikivi et al. 2005; Fedorova 2014).

The study ZERO was concerned with discourse coherence effects on the choice of referential form, as explained in the Centering theory framework (Grosz, Joshi, and Weinstein 1995; Walker, Joshi, and Prince 1998). Thus, the main focus in this study was on computing the Centering-based transition type (CONTINUE, RETAIN, SMOOTH SHIFT and ROUGH SHIFT). In addition to this theoretical Centering-based variable, three grammatical variables were also coded: 1) grammatical case, 2) syntactic role,<sup>14</sup> and 3) clause type. The choice of these three variables was based on the scarce literature regarding the phenomenon of Estonian zero reference (or ellipsis), mainly as presented in Estonian academic syntax (Erelt 2017a; see also Erelt et al. 1993) and in Lindström (2001, 2005)

---

<sup>14</sup> In ZERO, the term ‘grammatical role’ was originally used, but all other studies included in this thesis adopt the term ‘syntactic role’ (Croft 2001). Grammatical role and syntactic role are treated as synonyms in this thesis. It is assumed that for the purposes here, this terminological difference has no meaningful effect. ‘Syntactic role’ is preferred because of the term’s transparency and direct reference to the syntactic structure.

In DETERMINERS and PRONOUNS, the number of different variables was increased from four to nine. Besides the three grammatical variables also included in study ZERO (case form, syntactic role, clause type), one grammatical variable (number), two semantic variables (animacy of the referent, other animate entities in the clause), and three pragmatic factors were added (mention number, referential distance, type of previous NP).

The same variables as in DETERMINERS were also tested in PRONOUNS, and in addition, one pragmatic variable was added: type of previous NP. This variable was only included in the study focusing on independent pronouns. Pronouns usually cannot be used for introducing a new referent into discourse for the first time, at least not in a coherent narrative. While all pronouns are expected to have a preceding referential NP in a discourse, this does not hold for all detNPs, especially for NPs with indefinite determiners. Thus, the variable ‘Type of previous NP’ is not applicable for the data in DETERMINERS.<sup>15</sup>

A different set of variables was tested in the study DEMONSTRATIVES: i) the semantic class of the noun, ii) the concreteness of the noun, and iii) the semantic type of the verb. This study was interested in the more specific traits that distinguish the uses of two similar referential devices, namely two constructions with a definite determiner. One construction has a demonstrative pronoun as a determiner, and the other has a demonstrative adverb as a determiner. The more general variables that help to explain the choice of detNPs were already tested in DETERMINERS. Thus, DEMONSTRATIVES tests such variables that have the potential to add some more specific knowledge about the referential choice in terms of two constructional variants.

All variables used in the four studies of this thesis, along with the descriptions, variable levels, and corresponding studies, are presented in Table 4. The notions ‘clause’ and ‘utterance’ coincide in their meaning in this thesis to a large extent. The reasons why two parallel notions are used are related, firstly, to the chosen framework (i.e., Centering) in the first study, ZERO. ‘Utterances’ are to be understood as units of analysis in Centering theory (Grosz et al. 1995; Walker et al. 1998; see also Taboada and Hadic Zabala 2008). ‘Clause’, on the other hand, is a formal and entrenched notion for a grammatical unit that is smaller than a sentence and larger than a phrase and that contains a verb form and its argument(s) (Crystal 2008: 78). Secondly, the notions ‘utterance’ and ‘clause’ are also related to the nature of the analysed data, so that the former relates to the spoken (narrative) data and the latter to the written (corpus) data.

---

<sup>15</sup> The preceding NP might, of course, influence whether the determiner occurs at all in an NP. Also, certain definite determiners might require a certain type of preceding NP. However, these questions are left for future research to tackle.



**Table 4.** List of all explanatory variables tested in the thesis

Variable type	Variable	Description	Levels	Studies
CT-bound variable	Transition type	The transition type between two utterances as postulated in the Centering theory	CONTINUE, RETAIN, SMOOTH SHIFT, ROUGH SHIFT	ZERO
	Syntactic role	Syntactic role of the NP in a clause	subject, object, genitive attribute, other	ZERO, DETERMINERS, PRONOUNS
	Case form	Grammatical case of the referential NP	nominative, genitive, partitive, illative, inessive, elative, allative, adessive, ablative, translative, terminative, essive, abessive, comitative	
	Clause type	Syntactic clause type of the utterance containing the referential NP	main, subordinate	
Grammatical variables	Number	Grammatical number of the referential NP	singular, plural	DETERMINERS, PRONOUNS
	Mention number	Number of mentions in a row of the reference chain for the particular referent	1 <sup>st</sup> , 2 <sup>nd</sup> , 3 <sup>rd</sup> , further (4 <sup>th</sup> or further mention), implicit (referent is already implicitly mentioned), other	DETERMINERS, PRONOUNS
	Referential distance	Referential distance (in utterances) from the preceding mention of the same referent	0 (same clause), 1 (preceding clause), 2 (2 clauses back), 3 (3 clauses back), further (4 or more clauses back), no (no previous mentions)	
	Type of previous NP	Previous NP referring to the same entity	fullNP, detNP, pronoun	PRONOUNS

Variable type	Variable	Description	Levels	Studies
Semantic variables	Animacy	Animacy of the referent	animate, inanimate	DETERMINERS, PRONOUNS
	Other animate referents in the clause	The presence of multiple animate entities in the utterance	yes, no, only inanimate entities	
	Semantic class of the noun	Semantic class as can be derived from the primary meaning of the noun in a detNP	location, non-location, mixed (can be used as location as well as non-location)	DEMONSTRATIVES
	Concreteness of the noun	The concreteness of the referent based on the primary prototypical meaning of the noun in a detNP	concrete, abstract	
	Motion verb	The presence of a motion verb in a clause. Clauses that contain a motion verb (e.g., <i>minema</i> 'go', <i>lendama</i> 'fly') are tagged for 'yes'. All other clauses are tagged for 'no'.	motion verb, non-motion verb, no verb in a clause	

In addition to linguistic factors, three sociolinguistic factors – age, gender, and education – were included in the analysis in DETERMINERS. The sociolinguistic factors were analysed on a par with linguistic factors because the use of determiners is assumed to be strongly influenced by individual preferences. That is, even when some linguistic features are found to affect the choice of determiners, variation across speakers is highly probable, because there are no explicit grammatical standards for determiner use in Estonian. Multifactorial non-parametric statistical techniques conveniently allow to integrate factors from different levels of language into one analysis (Scrivner and Díaz-Campos 2016; Tagliamonte and Baayen 2012).

### **4.3. Statistical methods used for analysing the data**

After the coding process, the statistical analysis was carried out on each individual dataset individually.

#### **4.3.1. Descriptive statistics**

In the first study, ZERO, a descriptive overview of the transition types and their relation to the referential form was given in the form of contingency tables. Similarly, the relationships between other grammatical variables and their relation to the referential form was given. To estimate the association between variables, chi-square tests were performed. This procedure allows a researcher to check whether the differences between observed frequencies and the expected frequencies are statistically significant (e.g., Conover 1999; Levshina 2015).

The results of chi-square tests in ZERO should be considered with reservations, however, since not all of the requirements of performing the test were fully met. First, the observations are not fully independent of each other, since one speaker produced several different referential NPs which are all included in the data. Second, the expected (and observed) frequencies in some cells are smaller than 5. Nonetheless, studies in the vein of Centering theory have deemed the application of chi-square beneficial (e.g., Di Eugenio 1998; Hoffmann 1998; Hurewitz 1998). Therefore, ZERO includes chi-squared analysis to enable cross-linguistic comparison of Estonian Centering analysis to CT studies focusing on other languages. In future research, however, regression analysis would be preferred for similar analyses, especially when the goal is to take into account all the factors simultaneously.

### 4.3.2. Trees and forests

The data in DETERMINERS, PRONOUNS, and DEMONSTRATIVES have a multivariate nature, and all coded variables represent categorical scales. Therefore, non-parametric methods are required for making reliable statistical claims about the interactions between variables in the data (Levshina 2017; Tagliamonte and Baayen 2012). Here, two closely interrelated statistical techniques were chosen: conditional random forests analysis (Breiman 2001; Strobl et al. 2008; Strobl, Malley, and Tutz 2009) and conditional inference recursive partitioning tree models (Hothorn et al. 2006). As aptly put by Janda (2013: 26), these methods can collectively be referred to under the term ‘tree and forest’. A concise example of the application of the method can be found in Levshina (2017: 291–300).

The conditional inference tree analysis is statistically built on testing the significance of binary recursive splits of the data (Hothorn et al. 2006). The interactions between explanatory variables are presented in the form of a dendrogram, which enables a straightforward interpretation on the patterns present in the data (Strobl, Malley, and Tutz 2009). However, the downside of conditional inference trees is their variability and relative instability (*Ibid.*). To resolve this disadvantage, conditional inference trees are often combined with the random forests method (Janda 2013; Levshina 2015). Random forests make predictions based on a larger set of constructed tree models by selecting the most important variables for the data based on the average of all constructed trees (Breiman 2001; Tagliamonte and Baayen 2012).

Conditional random forests and inference trees have recently become a popular tool among linguists. For example, many researchers have found these tools useful for explaining the variation between more or less similar constructions in language (e.g., Baayen et al. 2013; Janda 2013; Levshina 2017; Tagliamonte and Baayen 2012). Several studies regarding the Estonian language also exhibit the appropriateness and versatility of ‘tree and forest’ analysis (e.g., Lindström et al. 2018; Lindström and Vihman 2017; Pook 2019; Siiman 2018; Taremaa 2017).

For all studies included in this thesis, the open source statistical environment R (R Core Team 2020) was used to perform the statistical computations, mainly by the help of the package *party* (Hothorn et al. 2020). The results are summarised in the next chapter.

## 5. MAIN RESULTS AND DISCUSSION

In this chapter, the main findings of the individual studies and the theoretical suggestions of the thesis are presented and discussed. The chapter begins with a description and explanation of the referential system in Estonian (section 5.1). The patterns of individual devices are first described (sections 5.1.1 to 5.1.4), and the overall system is characterised as the final result (section 5.1.5). The chapter then continues with a section devoted to exploring the features of the Estonian referential system in a cross-linguistic perspective, and contrasting the Estonian system with the workings of referential devices in Finnish and Russian (section 5.2).

### 5.1. The system of Estonian referential devices

#### 5.1.1. Zero reference

The first study, ZERO, set off from the problem that the referential properties of two closely related constructional variants – zero reference vs. the short form of the overt personal pronoun (*ta*) – have not yet been sufficiently described in Estonian. The study applied the Centering theory (CT) framework (Grosz et al. 1995; Walker et al. 1998) to investigate whether there is a connection between the preferred transition type and the third person form used in an utterance. ZERO can be seen as an extension of the purely salience-based model of reference, since it is assumed that both the overt form and zero reference express an equal level of referent salience. The theoretical question, then, is whether a rigid model (such as CT) that accounts for discourse coherence by considering salience as the leading factor can offer an exhaustive explanation for referential choice.

The basic notion of CT is a discourse construct called the ‘center’, i.e., a semantic entity in an utterance that links the current utterance to other utterances in a discourse segment. Based on a comparison of center types in the current utterance and the utterance that precedes it, three types of shifts are proposed within the framework. When one particular entity is talked about and kept in focus across two utterances, this indicates a CONTINUE transition. When one entity has been the focus of attention, and then another entity also gains attention in the following utterance(s), this will be called a RETAIN transition. Two utterances are connected by a SHIFT when the entities at the center of attention are different across utterance boundaries. In essence, the CT framework, and thus the first study in the thesis, test whether (and how) the choice of a referential device affects discourse coherence within a segment. Discourse is perceived as more coherent when the same centers are maintained throughout the segment.

In the data of 16 narratives based on the Pear Film (Chafe 1980), there were 238 overt personal pronouns (example 16) and 279 occurrences of zero reference (example 17),<sup>16</sup> 417 of which were singular and 100 plural forms. While zero reference is more frequent than the overt pronoun in spoken narratives, the differences between the transition types of the two forms are not considerable (at least without taking other factors into account). As expected, CONTINUE is the predominant transition type for both forms (165 CONTINUES with an overt pronoun and 259 with zero reference). Another anticipated tendency that emerged was that with zero reference, transition types other than CONTINUE are rare (3 RETAINS and 15 SHIFTS appeared in the data, though). The variation in transition types was somewhat greater with the overt form *ta*, which occurred in 35 RETAIN and 20 SHIFT transitions as well. In addition, in 18 utterances with the overt pronoun, it was not possible to determine the transition type, as compared to only 2 such utterances with zero reference.

- (16) *s= ta vist tahtis alguses*  
 then 3SG.SHORT.NOM probably want.PST.3SG beginning.INE  
*võtta ühte pirni=*  
 take.INF one.PRT pear.PRT  
 ‘then **he** probably at first wanted to take one pear’
- (17) *sis= ø sõitis seal mingi kruusatee peal*  
 then ø ride. PST.3SG there some gravel road.GEN on  
 ‘then rode there on some gravel road’

As the Centering analysis revealed only subtle differences in the usage patterns of the two forms, there still remains a large overlapping area in the usage patterns of *ta* and zero reference. Therefore, the effect of three additional factors (syntactic role, case, and clause type) was observed (see Table 5). It was found that there is a clear connection between zero reference and grammatical subjects in the nominative case. Out of 279 utterances with zero reference, 273 expressed the nominal subject in a clause. Only six utterances in the data deviated from this pattern and expressed non-subject in a clause. The overt form showed a bigger variation in possible syntactic roles and case forms. Nevertheless, subjects (139 out of 238) and nominative forms (135 out of 238) are also most frequent with the overt pronoun. The frequencies of the two forms in main and subordinate clauses were also calculated. In this regard, the overt form and zero reference proved quite similar in the data since both are usually used in main clauses. Only 38 overt forms and 5 zero forms came up in subordinate clauses (cf. 200 and 274, respectively, in main clauses).

---

<sup>16</sup> Examples 16 and 17 are adapted from study ZERO.

**Table 5.** Summary of the exploratory results of ZERO

Form	Transition type	Syntactic role	Case	Clause type
Column %				
Overt form <i>ta</i> (n = 238)	CONTINUE: 69.3	Subject: 58.4	Nominative: 60.9	Main: 84.0
	RETAIN: 14.7	Other: 41.6	Other grammatical	Subordinate: 16.0
	SHIFT: 8.4		case: 6.7	
	No Cb: 7.6		Local case: 30.7	
Zero reference (n = 279)	CONTINUE: 92.8	Subject: 97.8	Nominative: 97.8	Main: 98.2
	RETAIN: 1.1	Other: 2.2	Other grammatical	Subordinate: 1.8
	SHIFT: 5.4		case: 0.4	
	No Cb: 0.7		Local case: 1.8	
			Other adverbial case: 0	

The divergence in the referential properties of zero reference vs. overt third person pronoun connects to the third research question of this thesis, namely the issue of constructional variation of referential devices with seemingly similar functions in Estonian. The problem arises acutely when we consider the Givenness Hierarchy (Gundel et al. 1993) and the placement of a language's referential devices along the statuses proposed in this hierarchy. As the inventory of possible referential devices in a language is probably larger than six (the number of statuses proposed in the hierarchy), the question is how to further distinguish between the forms that fall under the same status, such as the Estonian personal pronoun and zero reference (for similar observations regarding English, see Scott 2013).

One possibility would be to explain the difference in terms of the meaning of the devices by saying that the overt form has a somewhat more specific meaning than zero reference. However, this is not the case, since both forms are equally vague and lexically underspecified (Garnham 2001), and thus, it is difficult, if not impossible, to distinguish between the forms according to meaning. One can just accept that (reduced) expressions that are similar in their referential properties probably share their meaning as well. However, the latter claim contradicts the principles of construction-based cognitive approaches in linguistics, which take as a premise that if two constructions differ in their syntactic content, their meaning and function, or their constructional profile must also differ (Goldberg 1995, 2006; Janda and Solovyev 2009).

Therefore, it is necessary to examine additional factors, pragmatic as well as grammatical, to determine the varying functions of different referential constructions. In this thesis, it is shown that Estonian zero reference is subject to certain grammatical constraints that do not affect the overt pronoun form. Namely, zero reference can usually be used with referents that occupy the nominal subject position in the main clause. If these conditions are not met, chances are that an overt pronoun is used instead (although zero reference is not

entirely infeasible). In addition, the zero reference was more constrained in terms of discourse coherence. So zero reference in Estonian can, and indeed has been, characterized as zero subject (Lindström 2005) and topic-drop (Kivik 2010), bearing some similarities to the English diary-drop phenomenon (Haegeman 2007; Scott 2013). Still, the more specified notion alone does not lead us closer to the reasons why the speaker chooses to omit the overt pronoun.

All things considered, the study ZERO suggests that it is not effective to explain the choice between the zero reference and overt pronoun forms merely based on the concept of discourse coherence or salience. For this reason, the rest of the studies in this thesis did not explicitly test the salience of the referent, but instead a combination of other factors was applied to see the referential patterns in the data. In PRONOUNS, the set of these factors is also applied to analyse zero reference as compared to other pronominal forms.

Research in progress has shed light on some further properties that can differentiate the overt form from zero reference. Namely, it seems that the overt form is preferred in contexts in which the speaker believes a sentence boundary (a full stop) to fall between two mentions of the referent. When no such boundary is assumed, i.e., when the speaker thinks that two subsequent mentions of the same referent belong to one sentence (e.g., are separated by a comma), then zero reference is preferred. Interestingly, the zero reference seems to be more sensitive to such sentential configuration than the overt pronoun (Hint and Kaiser *in prep.*). In addition, PRONOUNS indicates that referential distance influences overt personal pronouns. However, even when certain factors are more successful than others in predicting the choice of a particular form, it must be acknowledged that referential choice is not a fully categorical process and that a certain amount of underspecification and individual variation always remains in predicting models (Gundel et al. 2012; Hendriks 2016; Kibrik 1996; Kibrik et al. 2016). Therefore, whichever model or framework is chosen as a means to explain referential choice between several options (e.g., such as CT in ZERO), it is likely that some aspects still remain open.

### 5.1.2. Article-like determiners

The second study, DETERMINERS, was designed to systematically examine the use of referential NPs with article-like determiners in Estonian spoken narratives and to compare them with similar devices in Finnish. In this study, a multi-factorial approach was applied to similarly collected and coded data from two closely related languages. The first aim of this study was to comparatively investigate which definite and indefinite determiners are used and to what extent in Estonian and Finnish data and whether these determiners are grammatically similar or different. The second aim was to determine whether the same variables predict the use of a particular detNP in both languages or if there are device-specific differences across languages.



The analysis included 676 Estonian and 680 Finnish NPs that may occur in a syntactic position where the use of a determiner is possible. Out of these NPs, a determiner was used in 188 (27.8 %) Estonian NPs and in 121 (17.8 %) Finnish NPs. This result suggests that overall, determiners are somewhat more frequent in Estonian than in Finnish. Corresponding data from Russian was also collected, and certain pronominals, e.g., demonstrative pronouns or the numeral *odin* ‘one’, occurred in the determiner position in the data. However, such uses make up less than 10% of all Russian referential devices in our data, and therefore, Russian article-like determiners were not analysed in DETERMINERS.<sup>17</sup>

Turning now to the pronouns that most often function in the position of a determiner, it can be observed that the four most frequent determiners in the Estonian data are: i) the demonstrative pronoun *see* ‘this/the’ (example 18), ii) the indefinite pronoun *üks* ‘one’ (example 19), iii) the (possessive) pronoun *oma* ‘one’s’ (example 20), and iv) the indefinite pronoun *mingi* ‘some’ (example 21).<sup>18</sup>

- (18) *ning*      *aitas*                      *ä*              *selle*              *lohe*              *puu*  
 and      help.PST.3SG              umm              DEM.GEN              kite.GEN              tree.GEN  
*otsast*                      *alla*  
 from upon              down  
 ‘and helped umm **the kite** down from the tree’

- (19) *ää*              *üks*              *noormees*                      *sõidab*              *jalgrattaga*  
 umm      one.NOM              young man.NOM              ride.PRS.3SG              bike.COM  
 ‘umm **a young man** rides a bike’

- (20) *ta*                                      *sai*                                      *oma*              *tuulelohe*              *tagasi*  
 3SG.SHORT.NOM              get.PST.3SG              his              kite.GEN              back  
 ‘he got **his kite** back’

- (21) *mingi*                      *tüüp*                      *lennutab*              *lohet*  
 some.NOM              guy.NOM              fly.PRS.3SG              kite.PRT  
 ‘**some guy** flies a kite’

The ‘tree and forest’ analysis of Estonian data revealed that the most important linguistic variables that predict the use of a detNP are: i) mention number, ii) syntactic role, iii) referential distance, iv) animacy, and v) clause type. In addition, two social factors, age and gender, turned out to be significant predictors of determiner choice in Estonian data. The effect of age might be especially relevant for explaining the data from the viewpoint of grammaticalization, since older speakers might have more conservative preferences, reflected in the lower use of determiners. This hypothesis is even more supported by Finnish

<sup>17</sup> See PRONOUNS for a quantitative overview of Russian data and for examples.

<sup>18</sup> Examples 18, 19, and 20 are adapted from DETERMINERS. Example 21 is from the data.

data that also showed participant age as being the most significant factor. Notably, in Finnish data, age was relatively more important than in Estonian data. This can be related to the circumstance that in DETERMINERS, the Finnish speakers were older on average than the Estonian speakers, and as such, the higher importance of age and lower usage of determiners overall in the Finnish data might be a sign of time-related changes in preferences.<sup>19</sup> However, the more specific effect of sociolinguistic factors, especially age, on referential choice is left for future studies to discuss, since age-related usage patterns of determiners were not outlined in inference tree graphs in this study.

Three out of the significant linguistic variables (mention number, syntactic role and animacy) are also captured in the conditional inference tree dendrogram, based on which, three observations can be made in relation to Estonian detNPs.<sup>20</sup> i) When a referent is mentioned for the first time in discourse (i.e., indefinite reference), then *üks*NPs are mostly used for referring to animate entities, whereas *üks*NPs and *mingi*NPs are used to an equal extent for referring to first mentioned inanimate entities. ii) When a subsequent mention of the referent is made (i.e., definite reference), then in cases of subjects and genitive attributes, only *see*NPs are used. iii) Definite reference to objects and other syntactic roles is conveyed either with *see*NPs or *oma*NPs, or with NPs that entail some less frequent determiners. It must be kept in mind, however, that in each of the described groups, NPs without a determiner prevail. Still, in one group, the difference in the numbers of NPs without a determiner as compared to detNPs is marginal, namely in cases of indefinite mentions of animate referents.

The usage patterns of Estonian determiners fit well into the framework of grammaticalization with respect to their gradual diachronic change from more lexical items into less lexical, i.e., more grammatical elements (Heine, Claudi, and Hünemeyer 1991; Heine and Kuteva 2006). Based on many different languages, it has been shown that the numeral *one* is often the source of an indefinite article (e.g., Givón 1981; C. Lyons 1999; Weiss 2004), and anaphoric demonstrative pronouns are common ancestors of definite articles (Greenberg 1978; Diessel 1999; Heine and Kuteva 2006). The Estonian results conform to both of these quasi-universals. However, as can be gathered from the prevalence of NPs without a determiner, none of the observed determiners bear mandatory grammatical function. Therefore, determiners are not fully grammatical items and there is no reason to apply the grammatical label ‘article’ to them. Also, it is not possible to make reliable inferences about the stage of article grammaticalization in Estonian as compared to Finnish, since more data from different registers and usage contexts are required for that, as well as a different methodological approach.

The usage patterns of the indefinite determiner *üks* ‘one’ are a telling illustration that Estonian determiners are not full articles yet. As a determiner, *üks* exhibits an asymmetry in its uses with animate as contrasted to inanimate

---

<sup>19</sup> See Laury (1997) for a thorough discussion about the grammaticalization of articles in Finnish.

<sup>20</sup> For more details, see Figure 1 in DETERMINERS.

referents. This indicates that at the current stage, *üks* ‘one’ is still more sensitive to semantic selection criteria than grammatical ones. The preference for animate entities harmonizes with the observation that indefinite determiners in their initial stage of grammaticalization tend to be used when referring to the main characters, as discussed in Heine and Kuteva (2006: 105). Furthermore, the Estonian determiner *üks* is seen as the typical outset construction of a narration (Pajusalu 2009). On the other hand, indefinite *üks*NPs are not restricted to narrative-initial animate main characters, but they are used in other contexts as well. In principle, any new referent, regardless of its animacy, can be referred to with an *üks*NP, either at the beginning of a narrative or later on. While such uses were less frequent in the present data, they indicate the potential for the indefinite determiner *üks* ‘one’ to develop into a more grammatical device in Estonian.

Even more article-like than *üks* in Estonian is the definite determiner *see* ‘this/the’. First, *see*NPs depend on the grammatical features of the referent/NP, namely the syntactic role of the NP in a clause. Out of all syntactic roles, objects seem to be evoked most often with a detNP, but determiner use is also noticeable with subjects and other roles. The distribution of determiner forms across the roles moves from a *see*-only preference among subjects towards a *see*-and-other-equally among non-subjects and non-objects. Such distribution demonstrates that *see*NPs are not restricted to certain functions, but are evenly distributed across the possible usage scope. Second, *see* ‘this/the’ as one the most frequent determiners in Estonian (see also Pajusalu 1997c; Kirsipuu 2012), seems to be more available to the speakers than other determiners. The evidence that *see* is the best candidate for a grammatical article in Estonian also emerges from Taremaa et al. (2021), where the determiner *see* is compared to the determiner *too* ‘that’, and to adverbial determiners *siin* ‘here’ and *seal* ‘there’. It can be inferred from their study that *see* as a determiner is not biased towards any particular variable and can be used relatively freely in every observed context.

Besides *üks* ‘one’ and *see* ‘this’, other pronominals can also function as determiners, e.g., indefinite *mingi* ‘some’, possessive *oma* ‘one’s’, adverbials *siin* ‘here’, *seal* ‘there’, etc. As these determiners occur less often, however, it is not possible to make any far-ranging conclusions about their exact behaviour. Nevertheless, it is safe to say, based on their smaller frequencies and more specific usage contexts, that other determiners do not exhibit clear signs of grammaticalization into articles and on the cline of determiners to articles, they still fall close to the determiner-side.

### 5.1.3. Anaphoric pronouns

The third study, PRONOUNS, is based on the same methodological and theoretical foundations as DETERMINERS. The purpose of PRONOUNS was to reveal the usage-patterns of Estonian reduced referential devices in natural discourse and to compare these patterns to those displayed in Finnish and Russian data. The

study continued the analysis of the spoken narrative data that was also used in DETERMINERS, and thus variables included in the analysis are the same. Only one variable, the previous referential NP, was added.

As the data are the same as in DETERMINERS, the overall proportions of different referential devices remain consistent across the two studies, so that pronouns fall in between fullNPs without determiners as the most frequent referential devices and detNPs as the least frequent referential devices. 38.6% (n = 504) of all referential devices (n = 1304) were pronouns in the Estonian data (compare to 44% for fullNPs without a determiner and 17.3% for detNPs). The most often used reduced referential devices in Estonian are personal pronouns (52.8% of all pronouns; example 22, see also example 16 in section 5.1.1), followed by zero reference (37.1%; example 23, see also example 17 in section 5.1.1) and demonstrative pronouns (8.5%; example 24). Demonstrative adverbs as referential devices are used less often (1.6%; example 25)<sup>21</sup>, but this is also anticipated due to their specialised meaning that can express only spatial relations (in the picture-narratives, the focus was on the referents and the actions that they performed, not on spatial relations).

- (22) personal pronoun

<i>ning</i>	<i>ää</i>	<i>loheomanik</i>	<i>oli</i>	<i>õnnelik</i>	<i>et</i>
and	umm	kite owner.NOM	be.PST.3SG	happy.NOM	that
<i>ta</i>		<i>selle</i>	<i>lohe</i>	<i>kätte</i>	<i>sai</i>
3SG.SHORT.NOM		that.GEN	kite.GEN	into hand	get.PST.3SG

‘... and umm the kite-owner was happy that **he** got the kite back.’

- (23) zero reference

<i>jaa</i>	<i>sis</i>	<i>oli</i>	<i>poiss</i>	<i>jälle</i>	<i>rõõmus</i>
and	then	be.PST.3SG	boy.NOM	again	happy.NOM
<i>ja</i>	<i>õ</i>	<i>sõitis</i>	<i>rattaga</i>	<i>edasi</i>	
and	ø	ride.PST.3SG	bike.COM	further	

‘... and then the boy was happy again and **ø** rode further with the bike’

- (24) demonstrative pronoun

<i>kahjuks</i>	<i>jääb</i>	<i>see</i>	<i>õ</i>	<i>puu</i>	<i>otsa</i>	<i>kinni</i>
unfortunately	stay.PRS.3SG	this.NOM	umm	tree.GEN	upon	stuck

‘unfortunately **it** gets stuck at the top of a tree’

- (25) demonstrative adverb

<i>ja=</i>	<i>seda</i>	<i>sealt</i>	<i>kätte</i>	<i>saada</i>	<i>oli</i>
and	this.PRT	from there	into hand	get.INF	be.PST.3SG
<i>praktiliselt</i>	<i>võimatu</i>				
practically	impossible				

‘and it was practically impossible to get it back **from there**’

<sup>21</sup> Examples 22 and 23 are adapted from PRONOUNS, examples 24 and 25 are from the data.

The statistical analysis presented in PRONOUNS confirmed that in Estonian, six different factors are significant for referential choice: i) clause type, ii) syntactic role, iii) case, iv) referential distance, v) the presence of multiple animate entities in a clause, and vi) animacy (of the referent). Out of these variables, only animacy of the referent did not show up in the conditional inference tree dendrogram. Therefore, it was not possible to describe the more specific effect of animacy for pronoun choice.

The specific effects of these variables are as follows. First, demonstrative adverbs clearly stand out as a separate group because their use is strongly restricted by the case of the NP (only local cases). However, as there were only eight instances of independent demonstrative adverbs in the data, it is possible that in a dataset with larger number of demonstrative adverbs, some other influential factors also come into play. Second, the next influential variable was ‘other animate referents in the clause’, which helps to explain the use of demonstrative pronouns: when only inanimate entities are referred to, then demonstrative pronouns are usually used for subsequent mentions. When, on the other hand, animate entities are referred to, then mostly overt personal pronouns or zero reference are dominant. Third, in the case of main clauses, when referential distance is one clause, then zero reference is more frequent than the overt form. Longer referential distance is more likely to result in the overt personal form being used. The same holds for subordinate clauses: overt personal pronouns are much more often used there than zero reference. Note that this result is in accordance with the results obtained in ZERO.

Overall, the results from PRONOUNS show that the differences between personal pronouns and demonstrative pronouns, which have more distinguished referential properties in Estonian, are greater than those of personal pronouns and zero reference. Personal pronouns and zero reference often occur in similar linguistic environments, and their referential properties express more similarities than differences. However, the use of zero reference is more constrained in terms of grammatical as well as pragmatic factors, while the overt personal pronoun has no strong restrictions pertaining to its contexts. Thus, this thesis (especially ZERO and PRONOUNS) has shown that the third person pronoun *ta* ‘s/he’ is the most neutral reduced referential device in Estonian.

The frequency and usage patterns of pronouns as compared to other referential devices in spoken narratives seem to illustrate the relative importance of anaphoric reference in narrative discourse. Third person pronouns, both overt and zero reference, are substantially more common than demonstrative pronouns. The proportion of various pronominals is considerably different in the context of spatial (deictic) reference, for example, where demonstrative pronouns are the most frequent pronominals, while personal pronouns occur rarely and zero reference is negligible (Pajusalu et al. 2020; Reile et al. 2019). This variation falls well into the traditional distinction of anaphoric, i.e., in-text reference vs. deictic, i.e., in-space reference (Bühler 1990 [1934]; Diessel 1999; Halliday and Hasan 1976; Himmelmann 1996), according to which demonstratives serve as the markers of deictic referents, whereas personal pronouns are used for anaphoric

reference. However, this distinction is not categorical, as can be seen from the fact that quite often still, demonstrative pronouns function as anaphoric devices, and personal pronouns can be used in spatial settings (e.g., Reile et al. 2019). More recent theoretical approaches also stress the composite nature of certain referential expressions and concur that anaphoric and deictic reference are examples of the same cognitive process (Talmy 2017).

#### 5.1.4. Demonstratives

The last study, DEMONSTRATIVES, discussed the constructional alternation of two Estonian referential constructions with demonstratives as definite determiners. More specifically, demPronNP (example 26) and demAdvNP (example 27) constructions were compared. Differently from the other three studies in this thesis, DEMONSTRATIVES used corpus data instead of spoken narratives. Corpus data was chosen because, in spoken narratives, not enough demonstrative adverbs occurred in the determiner position, and thus, it was not possible to conduct a good comparison of these forms based on the narratives. However, demonstrative adverbs are productive in the determiner position in other contexts, since from the Estonian National Corpus 2017, 2400 clauses (1200 examples of both constructions) with demonstrative determiners could easily be collected. A semantic approach was taken, and the central question of the study was which semantic factors differentiate the use of a demonstrative pronoun from a demonstrative adverb if used in a determiner position in an NP. Three variables were tested: i) semantic class of the noun, ii) noun concreteness, and iii) verb type (motion verb or not).

##### (26) demPronNP

<i>Ohuline</i>	<i>on</i>	<i>see,</i>	<i>et</i>	<i>igaiüks</i>
important	be.PRS.3SG	DEM.NOM	that	everyone
<b><i>sellest</i></b>	<b><i>grupist</i></b>	<i>saaks</i>	<i>oma</i>	<i>teema</i>
DEM.ELA	group.ELA	get.COND	own	topic.GEN

‘It is important that everyone **from this group** would get her topic.’

##### (27) demAdvNP

<i>Aga</i>	<i>miks,</i>	<i>või</i>	<i>mis</i>	<i>seal</i>	<i>telemajas</i>
but	why	or	what	there.LOC	television house.INE
<i>üldse</i>	<i>toimub,</i>	<i>ei</i>	<i>saa</i>	<i>ma</i>	<i>aru</i>
whatsoever	go on.PRS.3SG	no	get.PRS.3SG	1SG.SHORT.NOM	mind.PRT

‘but why, or what goes on **there in the television house** [in there television house] whatsoever, I do not understand’

Two variables proved significant in the ‘tree and forest’ analysis: i) semantic class of the noun and ii) noun concreteness. Based on these variables, the

following patterns emerged from the data.<sup>22</sup> First, the analysis suggests that the choice between the two demonstrative determiners is semantically motivated. In corpus sentences, *demAdvNP* is used when the noun in the construction expresses spatial meanings, whereas *demPronNP* typically occurs when the noun clearly has non-spatial meanings (i.e., meanings related to time and animate entities). Importantly, in a group of nouns that express polysemous meaning (i.e., words that can be interpreted as either spatial or non-spatial, e.g., *laud* ‘table’, *auto* ‘car’), both determiners arise with equal frequency. Second, noun concreteness also affects referential choice. When the noun has abstract meaning, then a demonstrative pronoun is more often used as a determiner. In case of concrete meanings, the chances are higher for a demonstrative adverb to function as a determiner.<sup>23</sup>

An important implication from the study is related to polysemous nouns. A further analysis was conducted with only NPs belonging to the ‘mixed’ category. For this analysis, the determiner preferences of semantic subclasses (institution/building, thing, event, text(ual environment), other) were inspected. Further proof was found for the importance of the semantic class of the noun, although very straightforward patterns did not emerge. This analysis is an indication that in certain cases, the choice of a determiner can facilitate the interpretation of the construction as a whole. For example, *laud* ‘table’ should probably be interpreted as a place in *demAdvNP* construction, but as a thing in *demPronNP* construction. Together, the results from DEMONSTRATIVES suggest that the choice of a determiner, even in seemingly very similar, nearly synonymous constructions, is guided by deep underlying factors.

Along these lines, the study DEMONSTRATIVES illustrates how meaning-related factors can be helpful in explaining the constructional variation of full referential devices (see also Taremaa et al. 2021). In this study, it was shown that the meaning of a noun that occurs as the head of the NP has an effect on the choice of a demonstrative form. More specifically, the semantic class of the noun and noun concreteness can be used to predict whether a demonstrative pronoun or a demonstrative adverb is chosen in a particular construction. The semantics of the word has also been shown to affect the choice between proximal and distal demonstratives (Rocca, Tylén, and Wallentin 2019; Rocca and Wallentin 2020). In addition, the neurological importance of the concreteness of a word, e.g., the connection between abstract concepts and certain mental and social experiences, has been found in attempts to build a comprehensive brain-based semantic theory (Binder et al. 2016).

---

<sup>22</sup> See Table 3, Table 4, and Figure 2 in DEMONSTRATIVES for more details.

<sup>23</sup> It is known from earlier studies already that grammatical patterns and the semantic class of the referent can be interrelated, for example that animate referents commonly take subject position in a clause and inanimate referents are more common in object position (Helasvuo 2020). In DEMONSTRATIVES, however, different types of patterns are revealed, since in the data, *detNPs* were mostly optional, not core arguments.

According to usage-based approaches to constructions, the demonstrative and the noun must share certain semantic features that enable the appearance of a construction (Goldberg 1995; Gries and Stefanowitsch 2004; Janda and Solovyev 2009). However, many nouns have polysemous meanings. Such nouns may combine with determiners with different meanings, depending on which meaning is intended at the current speech situation (e.g., *siin koolis* ‘in this school [here]’ vs. *selles koolis* ‘in the school’). The results of this thesis are especially useful for explaining such cases. Namely, the choice of a demonstrative itself can realize the referential potential in language, i.e., the conceptualization of the referent (e.g., whether ‘school’ should be conceptualized as a building or as an institution) depends on the choice of a determiner.

An important difference related to the data concerns the mode of the analysed language. In the studies ZERO, DETERMINERS, and PRONOUNS, spoken language is analysed, but DEMONSTRATIVES deals with written language. Furthermore, in the context of DEMONSTRATIVES, it must also be noted that both formal and informal texts are included in the Estonian National Corpus 2017. These differences must be clearly addressed when making generalizations about the overall system of referential devices (see Taremaa et al. 2021). Thus, based on this research, it is safe to say that demonstrative adverbs as determiners are more common in written language than in spoken narratives. Regardless on the mode of communication, the choice of a demonstrative adverb as a determiner is determined by the semantics of the head noun, and the choice of a determiner itself can be one means of creating a referent.

### 5.1.5. Outline of the Estonian referential system

This thesis set out to give a systematic description of core referential devices in Estonian informal discourse, both spoken and written. The previous studies about Estonian referential devices take a form-based approach (Pajusalu 1997c, 2000, 2005, 2009; Kaiser and Vihman 2010; Kaiser and Hiietam 2003). This thesis differs from these previous accounts since the focus was on the referential chains in discourse. This means that the thesis seeks to explain the alternation of the possible referential devices that can be used for referring to similar entities.

The thesis covers the following referential devices in Estonian: i) zero reference, ii) personal pronouns, iii) demonstrative pronouns, iv) demonstrative adverbs, v) NPs with a definite determiner, vi) NPs with an indefinite determiner, and vii) fullNPs and proper nouns without a determiner. Most of these devices are cross-linguistically well-attested (Ariel 1990; Gundel et al. 1993; Kibrik 2011), one exception being demonstrative adverbs, which are not often regarded as common referential devices (see Diessel 1999: 74 for the special function of demonstrative adverbs). Furthermore, the results of this thesis show that zero reference is a full-fledged member of the Estonian referential system, although it has not been regarded in that sense before.



This thesis also aimed to uncover the factors beyond salience that facilitate a more comprehensive explanation of referential choice in discourse. As it is not practical to explain referential choice solely based on the referent's saliency (Kibrik 2011; Kaiser and Trueswell 2008), a list of additional factors was assessed in this thesis to see how forms that are similar in terms of salience can be further distinguished in discourse. Importantly, it was not expected to see the effect of all inspected variables on each referential device, but instead, it was presumed that different forms are sensitive to different factors, or that they are sensitive to one factor but to a different degree. This expectation was confirmed by the statistical multifactorial analysis of the data.

Table 6 below summarises the findings from the thesis and presents the multifactorial form-specific scheme of the core referential devices of Estonian natural discourse. The system of common referential devices in Estonian, as presented here, proceeds along the level of salience of the referent. Starting from the most salient devices (to the left) and moving towards the least salient devices (to the right), the additional form-specific factors that proved significant in explaining certain aspects of referential choice in Estonian are presented in the vertical dimension (top row). Some specifying remarks regarding particular devices are presented in the bottom row. Finally, the 'default', i.e., the least restricted devices, are marked for the reduced referential devices and full referential NPs.

In the light of one of the underlying assumptions in the thesis – that salience is too monolithic as a feature to adequately explain all the variance in referential choice – it is necessary to emphasize that the importance of salience in referential choice cannot be rejected. Due to the cognitive underpinnings of the processes that make a referent salient, it is deemed a fundamental feature in all human languages (Talmy 2017). It is generally acknowledged that a reduced form requires the speaker to have given its referent more attention in discourse, and more attention leads to higher level of salience of the referent of the chosen form (e.g., Chafe 1994; Kibrik 2011). However, the salience account can be strengthened by adding additional linguistic and also form-specific features, as we witness in the case of Estonian zero reference and overt pronoun, for example. A multifactorial approach enables a researcher to pinpoint more delicate differences beyond salience between various referential forms. For example, the salience level can remain the same, but two forms might just require different grammatical or semantic contexts, as also demonstrated in ZERO and in DEMONSTRATIVES.

In addition, the authors of one particular salience scale, Givenness Hierarchy, also stress in later work (Gundel, Hedberg, and Zacharski 2012) that different statuses should not be seen as encoding only one specific degree of salience, and that statuses only encode procedural information, saying not much about the conceptual content of the form. Therefore, salience is considered as an inherent feature for all languages, but it demands language- and form-specific specification.



Table 6 does not aim to indicate that these and only these factors affect referential choice in Estonian. Indeed, when putting this research into a larger perspective and comparing to other multifactorial studies, it appears that many other factors that were not explicitly tested here can also affect referential choice (for a similar discussion, see Kehler and Rohde 2013; Kehler et al. 2008). These include, for example, context, speech setting, register or genre (Pajusalu et al. 2020), or sociolinguistic factors, such as age and gender (as indicated in DETERMINERS). In addition, extra-linguistic factors, such as task-based expectations, goals, or the visual environment can also affect referential choice. (Brown-Schmidt, Byron, and Tanenhaus 2005). Furthermore, referential choice may be determined by speaker-internal factors, for instance by speaker's increased cognitive load (Vogels, Krahmer, and Maes 2015). In explaining the choice of deictic (spatial) referential devices (with a focus on demonstratives), distance between the referent and the speaker, visual salience, and contrast have proven influential (Reile 2019).<sup>24</sup> Some of these factors are tackled in anaphora-studies as well, for example the role of contrastive focus in reference resolution in English (Kaiser 2011a) and in Dutch (Kaiser 2011b) has been shown to be important. Distance, on the other hand, seems to be a purely deictic feature (Coventry et al. 2008; Diessel 2013), but it is important to study the possible effect of distance on anaphoric reference in future studies. This new perspective would be especially relevant in the context of more contemporary accounts, such as Etelämäki (2009) and Talmy (2017), suggesting a holistic view on reference that cover the processes of anaphoric and deictic reference under one framework.

Unfortunately, the data obtained for this thesis do not allow us to make any far-reaching conclusions about the distinction between Estonian long and short personal pronoun forms. This is because there were very few long pronoun forms used in the data (nine occurrences in ZERO and eight<sup>25</sup> in PRONOUNS). The infrequency of long pronoun forms in Estonian data has recently been attested by Vihman and Walkden (2021), as well. Therefore, Table 6 only captures the distinctions between the short personal pronouns as compared to other referential devices. It can only be hypothesized that the lack of long pronoun forms in the data is determined by the nature and content of the elicited narrative setting. In a more elaborate storyline, where pragmatic necessity for topic shifts and contrast would arise, the use of long pronoun forms would probably be more frequent. However, given that the long form has been previously described as having pragmatically determined rather than specific functions (Pajusalu 2005; Kaiser 2010; Erelt et al. 1993), it is anticipated to emerge relatively less frequently in language than the short form. In future studies, then, it will be necessary to

---

<sup>24</sup> As the focus is on anaphoric devices in this thesis, the above schema (Table 6) is intended to represent only anaphoric use of referential devices. A thorough analysis of the behavior of different devices in deictic use is presented by Reile et al. (2019; see also Pajusalu et al. 2020).

<sup>25</sup> In PRONOUNS, 41 occurrences of long pronoun form are reported. Of these, 33 occurrences are in genitive case and mostly (30) function as genitive attributes. As the functions of genitive attribute are referentially different from the independent pronoun, these forms do not allow to address the referential difference between long and short pronoun forms.

analyse the relative frequency and usage contexts of long pronoun forms in corpus data and compare these features to the characteristics of short forms.

## **5.2. A cross-linguistic perspective on referential devices**

Two studies in this thesis, DETERMINERS and PRONOUNS, tackle problems related to the cross-linguistic comparison of referential devices in discourse. The use of referential devices in Estonian is contrasted with uses appearing in two other languages, Finnish (DETERMINERS and PRONOUNS) and Russian (PRONOUNS). Thus, this thesis addresses the premise that referential devices with similar formal labels (e.g., personal pronoun, zero reference, determiner, etc.) can have various functions and usage practices in different languages. This perspective is not reserved for reference-related phenomena but has more general implications in linguistic theory (cf. Haspelmath 2010).

The results from DETERMINERS and PRONOUNS indicate that, indeed, there are substantial differences in the use of referential devices between languages. First, devices vary in their overall frequencies (Table 7) in spoken narrative discourse. The relative order of the three groups of refNPs is the same in Estonian, Finnish, and Russian: full NPs without determiners are used most often, pronouns fall in between, and full NPs with determiners are used most sparingly. Closer inspection of the table, however, shows that in Estonian, pronouns are used slightly more often than in Finnish but less than in Russian discourse. The usage frequencies of NPs without determiners are similar in Estonian and Russian, whereas Finnish stands out in this regard, as more than half of the referential devices are fullNPs without determiners in the Finnish data. Full NPs with determiners occur most often in the Estonian data, slightly less in the Finnish data and are rare in the Russian data. Taken together, each language has a certain referential device for anaphoric reference that is “preferred” in comparison to other languages. In Estonian, this device is detNP, in Finnish, it is NP without determiners, and in Russian, pronouns stand out as being used to a larger extent than in other languages.

Turning now to the more fine-grained differences, Table 7 reveals interesting tendencies in pronoun use as well as in detNP use. As can be expected, personal pronouns are the most used pronouns in all three languages. What catches the eye, though, is the high selection of demonstrative pronouns in Finnish. As much as 25% of all pronouns are demonstrative pronouns in Finnish data (cf. 8.5% in Estonian and 1.5% in Russian). Russian, on the other hand, gives higher priority to zero reference (45.7%, as compared to 37.1% in Estonian and 36.6% in Finnish). Regarding NPs with determiners, it can be observed that demonstrative pronouns are the most frequent determiners in the sample of article-less languages studied here. However, variation can be seen in the demonstrative stems (one demonstrative in Estonian and Russian vs. two demonstratives in Finnish). In addition, Estonian seems to use indefinite determiners quite often, while this device is rare in the Finnish and Russian data. In sum, the languages

exhibit quite different patterns of pronoun usage, even when the inventory of referential devices is relatively similar.

**Table 7.** Summary of referential devices used in the data of three languages

	Estonian	Finnish	Russian
pron	38.6% (504)	31.8% (464)	<b>45.3% (530)</b>
	persPron 52.8% (266) demPron 8.5% (43) demAdv 1.6% (8) zero 37.1% (187)	persPron 35.1% (163) demPron 25% (116) demAdv 3.2% (15) zero 36.6% (170)	presPron 52.4% (278) demPron 1.5% (8) demAdv 0.4% (2) zero 45.7% (242)
NP w/o det	44% (574)	<b>53.8% (785)</b>	46.8% (548)
detNP (incl. genNPs with nominal and pronominal genAtt)	<b>17.3% (226)</b> genNP 17.2% (39) <i>see</i> NP 41.6% (94) <i>üks</i> NP 19.0% (43) otherNP 22.1% (50)	14.4% (211) genNP 31.3% (66) <i>se</i> NP 19.9% (42) <i>tämä</i> NP 17.5% (37) <i>yksi</i> NP 3.3% (7) otherNP 28.0% (59)	7.9% (93) genNP 19.4% (18) <i>eto</i> NP 43.0% (40) <i>svoj</i> NP 25.8% (24) otherNP 11.8% (11)
Total	100% (1304)	100% (1460)	100% (1171)

The raw frequencies per se cannot offer an explanation of the underlying reasons of why certain patterns arise in some languages but not in others. Therefore, it is necessary to examine the possible variables that have an effect on the choice of a linguistic device. For that purpose, this thesis looked at nine grammatical, semantic, and pragmatic variables. Based on the parallel ‘tree and forest’ analyses run with Estonian, Finnish, and Russian data, it can be seen that these variables have different effects in languages. Furthermore, different devices exhibit dissimilar responses to variables across languages.

Table 8 presents the summary of all the variables examined in DETERMINERS and PRONOUNS.<sup>26</sup> If, based on the statistical analysis, it was possible to clearly infer the influence of a certain variable on a particular referential device, then this device and its corresponding context are also reported in the table. From the table below it can be seen that both pragmatic and grammatical variables are important in all three languages. However, Estonian determiner use is explained by pragmatic factors, while in Finnish, grammatical factors seem slightly more influential. Estonian pronoun use, on the other hand, is primarily guided by grammatical factors, although pragmatic ones also have an effect. In Finnish and Russian, the impact of grammatical and pragmatic factors is more balanced.

<sup>26</sup> As the Russian data was poor regarding the use of detNPs, the significances of variables relevant to determiner use in Russian is not reported; also, the analysis of the Russian pronoun data was capable of explaining only the usage patterns of personal pronouns and zero reference.

Based on these results, the following theoretical suggestions can be pointed out. First, relating to the usage frequencies of different types of anaphoric expressions, it is suggested that each language can be characterized in terms of its “favourite” anaphoric device: that is, referring practices vary across languages (Hanks 1990). In this respect, Finnish can be considered a demonstrative-biased language and Russian is a personal pronoun-biased language. Estonian falls somewhere between these types. A similar observation was made in a study that investigated the referential properties of relative clauses in Estonian, Finnish, and Russian, where it came out that Finnish speakers prefer to use demonstrative pronouns, but Russian speakers use more relative clauses (Pajusalu et al. 2018).

As there are three demonstrative pronouns in Finnish, two in Russian, and two or one in Estonian, it can be tempting to claim that the number of demonstrative stems in a language is the reason for preference patterns. However, the number of stems in itself does not explain why demonstratives are used more in one language than in other. After all, the “purely” demonstrative functions should be expressed to a similar extent in languages, regardless of the particular linguistic expressions that embodies this function. Instead, the usage frequency is related to the functions of demonstrative pronouns in a language. For example, in Finnish, the demonstrative *se* ‘it/this’ regularly refers to human referents in discourse and is even considered to be a personal pronoun by some researchers (see also Hakulinen 1985). Thus, in Finnish, demonstratives are often used in contexts that would require some other device (personal pronoun or zero reference) in Estonian and Russian. Therefore, even when the label ‘demonstrative’ or ‘personal’ pronoun is assigned to a device, this does not imply that these are used in an analogous way across languages.

The second aspect of cross-linguistic variation relates to the usage frequencies of different determiners. Again, the data revealed quite different usage of article-like determiners in languages, although the grammatical structure of Estonian, Finnish, and Russian allow the use of determiner NPs. The process of grammaticalization may be seen as responsible for such differences. As explained in Heine and Kuteva (2006), the process of grammaticalization has different stages and, respectively, the languages may be on different stages of this process. Furthermore, grammaticalization is not a “must” – a language may not be sensitive to this change at all, as the Russian data indicates. This idea is elaborated further in DETERMINERS.

Taken together, these results suggest that referential forms are sensitive to different factors in different languages and that languages exhibit important differences in their referential practices. The differences in this regard may also pertain to the complex issue of language contacts. However, more thorough analysis of this issue falls out of the scope of this thesis.

**Table 8.** Factors that help to explain referential choice in Estonian, Finnish, and Russian. + = significant explanatory variable according to both inference tree and random forest analysis; (+) = significant variable in random forest but not in inference tree analysis; \* = variable that showed in the inference tree dendrogram, but proved not to be significant in the random forest analysis.

		Estonian	Finnish	Russian
<b>Determiner use</b>				
<i>Semantic factors</i>	Animacy	+	*	
		(indefNP)	(defNP)	
	OtherAnim			
<i>Pragmatic factors</i>	MentionNo	+	+	
	RefDist	(+)		
<i>Grammatical factors</i>	Number			
	Case		+	
			(demAdv+NP; inanimate defNP)	
	SyntRole	+	+	
		(defNP)		
	Clause type	(+)		
<b>Pronoun use</b>				
<i>Semantic factors</i>	Animacy	(+)	(+)	
	OtherAnim	+	+	
		(Inanimate only: demPron)	(inanimate only: demPron)	
<i>Pragmatic factors</i>	MentionNo	*	(+)	
	RefDist	+	+	+
			(≥1: Zero; ≤2: persPron)	(1: Zero; ≠1: persPron)
	PrevRefNP			
<i>Grammatical factors</i>	Number			
	Case	+	+	+
		(spatial: Adv)	(ABL, ILL: Adv)	(non-nom: persPron)
	SyntRole	+	+	(+)
		(animate non- subject: PersPron)	(non-subject: PersPron)	
	Clause type	+		
		(subject, subordinate: persPron; Subject, main: Zero)		

## **6. METHODOLOGICAL IMPLICATIONS AND LIMITATIONS OF DATA COLLECTION AND ANALYSIS**

In addition to the theoretical facets of the study of reference, the thesis has also raised some methodological issues regarding the inquiry of referential choice. The methodological aims relate to i) the narrative-elicitation method, ii) the appropriateness of corpus data for investigating referential devices, iii) eliciting or obtaining the infrequent or ‘hidden’ referential forms and constructions for looking into the constructional variation of referential devices, and iv) the applicability of Centering theory in analysing Estonian data<sup>27</sup> (see also Chapter 1). The sections to follow in this chapter will discuss these issues along two sets of questions: i) questions related to data collection (section 6.1), and ii) questions related to data coding (section 6.2). In this respect, this thesis looks at the advantages and drawbacks of spoken discourse as set against the written corpus data for studying referential forms and the options for extracting less-frequent referential forms for quantitative analyses.

### **6.1. Collecting data for finding referential devices in discourse<sup>28</sup>**

#### **6.1.1. Narrative data**

Narratives have proven to be a useful data type for reference studies, including both spoken (elicited) narratives (e.g., Chafe 1980; Du Bois 1987; Berman and Slobin 1994; Hickmann 2004; Hedberg 2010; Serratrice 2013; Disbray 2016; Verhagen 2019; Dancygier 2019) as well as written prose (e.g., Kibrik 1996; Krasavina 2011; Arnold 2003; Kaiser 2015). Three studies in this thesis were also conducted based on spoken narrative data. Importantly, these narratives are quasi-experimentally elicited, which means that besides the conveniences of narrative data, certain hindrances must also be taken into account (see also Pavlenko 2008).

First, as in every study that puts an emphasis on human participants, the instructions of the experimental setting can have a considerable effect on the participant’s behaviour and the language produced. When the instructions leave too much room for personal interpretation for the participant, she might produce

---

<sup>27</sup> Centering theory and its advantages and disadvantages in analysing Estonian third person reference are discussed in section 5.1.1.

<sup>28</sup> In addition to (elicited) narrative data and corpus data, another widely used method for collecting linguistic data is (psycholinguistic) experiments (for an overview, see Gonzalez-Marquez et al. 2007). Experiments are also very useful for reference studies (e.g., Arnold 2001; Arnold and Griffin 2007; Fukumura and van Gompel 2015; Kaiser 2011a; Kaiser and Trueswell 2008; Kehler and Rohde 2013; Vogels et al. 2015, just to name a few). In this thesis, however, linguistic experiments are not discussed in more detail.



the kind of data that would be not entirely relevant to the study. For example, the participant can choose to produce a static description of the pictures without combining the depicted events into a coherent and consistent storyline (e.g., *here's a boy riding a bike. now here the boy runs over a big stone. oh no, now the boy cries*). A description of this type would yield a text that has few benefits for the study of reference. Second, the principles of referential choice would not emerge in a narrative in which a participant gives proper names to the depicted characters and only uses the names without switching these to other referential devices (e.g., *Peter is riding a bike. Suddenly, Peter runs over a big stone. Peter starts to cry*). Therefore, the instructions for the narrative task must inform the participant very clearly from which perspective they should approach the narrative and to whom they should direct it (see also Pavlenko 2008).

On the other hand, too-detailed instruction could also have a negative effect on data naturalness and quality. Thus, one can argue that a better, available alternative to (elicited) spoken narratives would be spontaneous speech data, e.g., conversations (as used by Helasvuo 2003, 2001, 2020; Kärkkäinen 1996, for example) in which the researcher does not have to account for possible issues of data not being natural or authentic. Spontaneous speech data certainly has its advantages when particular referential forms and their frequency are considered, as numerous studies can confirm (e.g., Monzoni and Laury 2015; Gundel and Johnson 2013; Huang 2000; Etelämäki 2009). Nevertheless, if the aim is to study the behavior of different referential devices in continuous reference chains uninterrupted by other speakers, then spoken conversations probably do not provide the kind of input that meets these requirements. On one hand, the conversational data does not provide a collection of narrative stories which are spoken by different speakers and in different languages, but on similar topic and about similar referents. On the other hand, difficulties may arise due to the nature of spoken conversations, which include turn-taking, reductions, hesitations, (self-)corrections, discontinuous talk, etc. (Sacks, Schegloff, and Jefferson 1974), and in which the topics and referents alter in a more hectic manner. Many of these features, of course, also occur in elicited spoken narratives, but these are easier to spot and analyse in speech produced by only one speaker and in a clearly restricted context.

Having discussed the benefits of narrative data, there are several options and settings available for collecting narrative data under relatively controlled conditions. This thesis reports two different narrative elicitation tasks. In ZERO, a well-established and well-known narrative elicitation setting from the Pear Stories project (Chafe 1980; Du Bois 1987) was used. In this study, the main concern was the naturalness of the narratives. For this reason, the experimental task was modified so that the narration was arranged between pairs of close friends. It was the first requirement in the instructions that the participant must come to the experiment together with a friend of her own choice. The participant watched the Pear Film and the friend waited in another room. After seeing the film, the participant had to relate to her friend what had happened in the film. This arrangement ensured that i) the speaker knows the hearer well and

feels comfortable narrating and ii) the speaker performs the narration naturally, since she knows that the hearer has not seen the film and she communicates directly to the addressee. Thus, this task was expected to yield data as close to that of real-life interaction as possible.

In DETERMINERS and PRONOUNS, a different task was used; namely, narratives were elicited by picture-sequences. The elicitation sequence was mainly inspired by Koster et al. 2011), a study about subject pronouns in child Dutch (but see also Disbray 2016). The main motivation was the simplicity of the task. The aim was to apply a simple and straightforward elicitation task to collect controlled and comparable data across many languages. For applicability to cross-linguistic field-work, the task was designed to be relatively quick and applicable in various conditions (e.g., in indoors and outdoors, with young and elderly participants, in places without technical devices and internet connection, etc.).

While the picture-sequence-based elicitation is indeed easy to carry out, it comes with its pitfalls, because it is generally known that the researcher herself has an influence on data that is produced via elicitation (e.g., Lehmann 2004). Specific to the picture-sequence-based task used for this thesis, two main aspects deserve attention. First, the participant can feel uncomfortable when telling a story to a stranger (i.e., to the researcher) and this definitely affects the naturalness and overall quality of the language. Second, the task itself might feel awkward or silly to (adult) participants, which also leaves traces on their produced language. For example, the aforementioned strategy to mention the story characters by proper names only can be a consequence of such discomfort or uncertainty. Also, participants can produce very short narratives (two or three clauses) under the stress of the experiment. As the data obtained for this thesis shows, most participants are collaborative, and the data quality is suitable for studying referential choice and the alternation of different devices in discourse. This is reflected in the fact that from the final data it was not necessary to exclude any narratives that exemplified the higher than usual stress level of the participant. However, some narratives that did not correspond to the instructions (e.g., they were descriptions rather than narratives) had to be left out from the analysis.

### **6.1.2. Corpus data**

Narratives and spoken discourse in general are not, of course, the only possible data type that is useful for studying referential devices. Many researchers describe their studies as based on corpus data (e.g., as in Kaiser 2015; Kibrik 1996; Krasavina 2011). However, there are different possible interpretations of what corpus data can mean, which also has its implications for the nature and analysis of the data. Among many researchers, the term ‘corpus’ is used in its wider sense as referring just to a collection of text(s), either in written or spoken form. Such uses of the term do not make any assumptions about the methodology that

is applied for analysing the text obtained from a corpus. On the other end of the spectrum stand linguists who consider only digital machine-readable, balanced and representative collections of texts as a proper corpus, and the nature of such corpora implies certain methodological consequences (for an overview, see Tognini-Bonelli 2011; Gilquin and Gries 2009). This thesis expresses concordance with this latter, more restricted conception of corpus data.

The application of quantitative corpus data as used in DEMONSTRATIVES does not come without its limitations, of course. Corpus data is often limited since access to whole referential chains is impossible. This means that in relation to referential choice, quantitative corpus data is useful in cases in which answering a research question does not require access to the complete discourse. For example, one can obtain an overview about the semantic or environmental (i.e., syntactic, morphological, phonological) characteristics of the word(s) accompanying the observable device. As such, corpus data is only useful for analysing the differences between linguistic expressions on a more local level. Thus, determining a referential device's relation to other referential devices is probably not achievable in corpus data, as it is not possible to observe how the devices alternate when the discourse unfolds.

The advantage of considering referential devices in overall discourse is that a researcher is able to follow the whole referential chain throughout the entire discourse and make reliable inferences about the referential functions of a particular device (e.g., Kibrik 2011). This claim can be understood such that the researcher should always have access to whole referential chains in discourse in order to make valid claims about the referential functions of a particular device. However, written corpus data has also proven useful in the study of referential choice (Kibrik et al. 2016). This thesis demonstrates that certain characteristics of referential devices can be determined based on individual clauses without or with little surrounding context, such as clauses obtained from corpus data. In this thesis, corpus data supplements full (narrative) discourse because certain referential expressions (here, NPs with a demonstrative adverb as a determiner) occurred rarely in Estonian narrative data. However, the mere fact that an expression occurs infrequently in a certain text type, genre, or context is not an indication of this device's lesser importance in the language system (Gries and Ellis 2015). For example, and as the corpus data in DEMONSTRATIVES shows, demonstrative adverbs are often found in a determiner position in a large body of natural, non-elicited language use (i.e., in corpus data). Therefore, the relevance of corpus data for the study of reference should not be neglected.

One important shortcoming of feature-based corpus analysis, however, is that it does not allow one to make inferences about the contexts in which one or the other device is implausible or even impossible. This weakness is also stressed in Taremaa et al. (2021), where it is concluded that probabilistic models of language (Bresnan 2007; Stefanowitsch 2008; Grafmiller 2014) obtained via corpus data are incapable of determining whether the linguistic constructions under study are largely interchangeable (i.e., fully probabilistic) or whether one of the constructions has certain highlights that do not allow this particular

construction to occur under certain constraints (Taremaa et al. 2021). Therefore, to fully understand the behaviour of the expressions under observation, it is advisable to combine corpus data with data containing full discourse and experimental data. This approach is, of course, not particularly novel in linguistics (see e.g., Gilquin and Gries 2009; Klavan 2012). However, future studies regarding specific referential devices and referential systems in general would also benefit from integrating information from different types of data. For example, the characteristics of the third person long personal pronoun (*tema*), as compared to the short form (*ta*), could be elaborated in greater detail in studies in which corpus data and a large number of clauses with *tema* are analysed.

## 6.2. Coding data for analysing referential devices

Two studies in this thesis, ZERO and DEMONSTRATIVES, focus on (referential) differences between a certain sets of referential devices that occur in seemingly similar referential settings and express similar meanings. In ZERO, the referential and contextual differences between the overt third person pronoun and zero reference were investigated. In DEMONSTRATIVES, two determiner constructions were compared, namely the demonstrative adverb vs. demonstrative pronoun in a determiner position in an NP. A unifying trait of the two studies is their attention to the question of how to determine the alternative functions and contexts of referential devices that are used in very similar conditions and have similar meaning, i.e., the question of constructional variation (Goldberg 1995; Gries and Stefanowitsch 2004; Silvennoinen 2018).<sup>29</sup> The methodological differences between these studies are mostly related to decisions associated with the data coding procedures. That is, in ZERO, ways to code an invisible i.e., phonologically unmarked device in the data are sought. In contrast, in DETERMINERS, the invisibility of one device shows in frequency effects: low frequency of a device in one type of language mode (i.e., in spoken narratives) guides the research to another type of language mode (i.e., written corpus), where this device occurs with considerably higher frequency.

### 6.2.1. Detecting and tagging for possible predictor variables

Many reduced referential devices seem to have only subtle semantic differences (consider, for example, the difference between zero reference vs. the overt 3SG

---

<sup>29</sup> An terminological clarification is in order here. Using the terms ‘near-synonyms’ or ‘constructional alternation’ is avoided in the thesis, since these are usually defined as regularly mutually interchangeable constructions (e.g., Arppe 2009: 13). As the referential constructions compared in this thesis are probably not always interchangeable (although in many contexts, they are), a broader term ‘constructional variation’, is deemed more suitable (see also the introduction in DEMONSTRATIVES and Taremaa et al. 2021).

pronoun, or pronominal vs. adverbial definite determiner in Estonian).<sup>30</sup> The circumstance that such referential devices often occur in grammatically similar linguistic environments can complicate telling these forms apart. However, when there is a visible difference in the level of form, a semantic difference is also expected to arise (e.g., Langacker 1987; Goldberg 1995). Therefore, it is crucial to determine the possible factors above or below the formal level that might influence the behaviour of alternative forms in discourse.

One study in this thesis, DEMONSTRATIVES, has applied the constructional perspective (Goldberg 1995) and adopted the alternation studies framework (e.g., Boas 2008; Silvennoinen 2018; Janda and Solovyev 2009) with a purpose to discover the differences in the referential properties of similar forms, that, according to the theory, bear different functions in discourse. Thus, the study DEMONSTRATIVES scrutinised several semantic properties of other words in a clause (nouns as well as verbs) that accompany the observed referential expression in a larger construction. That is, the noun's concreteness and semantic class were tagged and the semantic class of the verb was annotated. In the coding procedure, it became evident that a human tagger inevitably brings subjectivity and discrepancies into the interpretation of semantic factors. For example, it is possible to count the noun *voodi* 'bed' as a thing (furniture) or as a location (place). However, such inconsistencies can be reduced by exercising a multi-coder procedure. Taken together, results from DEMONSTRATIVES suggest that the semantic properties of the words that accompany a particular reduced device are useful predictors of that device's referential properties.

Methodologically, the number of variables is not usually limited in studies of referential devices. For example, Kibrik et al. (2016) have attempted to predict referential choice in discourse by incorporating as many as 25 factors into their model. Yet, Kibrik et al. (*Ibid.*) acknowledge that the model is open for additional factors to be included. However, analysing a larger set of variables requires applying machine learning methods and computational analysis. In the studies conducted for this thesis, only a limited number of variables were coded: three in DETERMINERS to nine in PRONOUNS. This limited number of variables is a result of manually tagging the data. In future investigations, it might be useful to use a more comprehensive set of variables.

When conducting a multi-factorial study and including several predictors into the analysis, a probable outcome is that several predictors will prove significant in a statistical analysis (e.g., Klavan 2012; Tagliamonte and Baayen 2012). The present thesis has also demonstrated that the choice of a referential device can depend on many factors, either grammatical, semantic or pragmatic, that exist simultaneously. A question about the relative importance or significance of the chosen factors thus arises. This thesis does not offer a comprehensive explanation about the relative importance of each factor on the individual choice of a

---

<sup>30</sup> By definition, reduced referential devices have little semantic content, i.e., meaning (Kibrik 2011: 39), therefore it is difficult to determine their semantic properties based on the reduced device alone.

particular device. While the random forest analysis output arranges the predictor variables according to their relative importance, it does not make any implications on the level of a particular device, but presents the results as a combination of the independent variables included in the analysis. For example, as in PRONOUNS, the influential factors for four pronoun types (zero, personal pronoun, demonstrative pronoun, demonstrative adverb) were combined and investigated under the category of ‘Phrase type’ as an independent variable. As an outcome, a unified list of significant factors does not signal exactly which device is sensitive to which factor. Thus, it is not discussed here to what extent each factor helps to explain the choice of a certain referential device (see DEMONSTRATIVES for a more detailed discussion of this question). Therefore, a welcome future development of this study would be to investigate the degree of importance of each predictor (e.g., by using logistic regression models) and to discover whether the most important predictor remains similar across devices, or if each device has its own most important predictor.

### 6.2.2. Finding the alternative variants

Some methodological questions arise in relation to detecting instances of zero reference in naturalistic discourse. It remains yet unresolved which overtly unmarked syntactic positions should and should not be considered as instances of zero reference. It is relatively straightforward to count phonetically unmarked subjects as occasions of zero reference, because most verbs need to combine with at least one argument function to form a clause. In addition, verb endings in affirmative forms also signal the person in Estonian (Erelt et al. 1993; Lindström et al. 2008). However, it is more complicated to decide whether there is a zero in clauses in which non-subject arguments, and maybe even non-arguments remain phonetically unexpressed. For example, in the sentence *sööb siis muheledes seda õuna* ‘then (ø) eats this apple, chuckling’, the expression of subject with zero reference is apparent, since there must be an actor who does the eating. However, in the sentence *sööb siis muheledes* ‘then (ø) eats (ø), chuckling’ it is difficult to decide whether there is a zero object or not, because with the verb *sööma* ‘eat’, transitive as well as intransitive uses are possible. Thus, a varying degree of transitivity in certain verbs poses a problem for coding possible zero objects.

In ZERO, an attempt was made to find subject as well as non-subject zeroes in the narrative texts. The coding of zero reference was one of the most complex step in the process of data treatment. Decisions had to be made about how to count the clauses with this referential device, and how to be able to quantify the findings and compare it to frequencies of other devices. Therefore, the analysis in ZERO is based on the verb forms that do not have explicit verb arguments in positions where one would expect to have a non-zero element (see also Kibrik 2011: 105). The final coding provided only six non-subject zeroes, as compared to 273 subject zeroes. Regarding these six instances of non-subject zeroes,

however, the zero-less interpretation is not completely unfeasible since the verbs in these clauses can be used either in transitive or intransitive constructions. The only exception was the verb *panema* ‘put’, which probably always requires a transitive interpretation by a native Estonian speaker, i.e., a clause with the verb *panema* should evoke both subject and object referents (example 28).

- (28)  $\emptyset$     *pani*=                     $\emptyset$  (?)    *endale*                    *sinna*    *ratta*                    *peale*  
 $\emptyset$     put.PST.3SG     $\emptyset$             oneself.ALL            there    bike.GEN            on  
‘ $\emptyset$  put  $\emptyset$  there onto his bike’

It was out of the scope of this thesis to offer straightforward operationalization principles for detecting zeroes in the non-subject position in a clause. Therefore, to avoid erratic coding of zero reference, only zero subjects are accounted for in PRONOUNS. This practice gains further support from cross-linguistic observations that zeroes tend to be more common with subjects than other arguments (Siewierska 2004). However, when a linguistic device is uncommon, it still requires deeper investigation in order to reveal all the workings of human language and cognitive mechanisms. Therefore, it is left for future studies to decide whether it is even possible to detect non-subject zeroes in natural language data. Moreover, from the quantitative perspective, the question remains whether zero reference can also be extracted from corpus data.

## 7. CONCLUSION

This thesis has discussed linguistic reference, a crucial part of language usage. The research was designed to examine the usage of the main referential devices in Estonian discourse. The main means to refer to entities around us are noun phrases. In this study, different types of referential noun phrases were observed, including full noun phrases (*poiss* ‘a boy’, *see ratas* ‘the/this bike’), demonstratives (*see* ‘this’, *too* ‘that’, *seal* ‘there’), pronouns (*ta* ‘he/she’) and zero reference. More specifically, various factors that guide the choice of a particular device at a particular position in the reference chain were investigated. The main theoretical tenet in this thesis derives from the form-specific multiple constraint approach (Kaiser and Trueswell 2008; Kibrik 2011), which claims that salience alone is not enough to reveal the more subtle influences behind the choice of a referential device. In addition, the analysis included the effect of an individual language, since it was presumed that in different languages, grammatically similar referential devices have diverse functions (see also Kaiser 2013).

The aim of the thesis was threefold. First, the thesis aimed to systematize the description of Estonian referential devices based on the features that most often characterize their usage in natural discourse. The second objective was to investigate structurally and semantically similar referential constructions that, despite of their seemingly comparable function, have quite dissimilar usage preferences. And third, this thesis intended to contrast the behaviour of the Estonian referential system with two other languages (Russian and Finnish) in order to pinpoint whether there are certain features or devices that differ in Estonian.

One of the most significant implications of this study is support for the argument that the system of Estonian referential devices (or more broadly, the system of referential devices in any language) cannot be interpreted as falling on a unified, salience-based scale. Instead, the use of referential devices depends on multiple factors working together simultaneously. These factors belong to different levels of language structure, such as i) grammatical level (case, number, syntactic role, clause type), ii) semantic level (semantic class, concreteness, animacy or other animate entities in a clause), or iii) pragmatic level (mention number, referential distance, type of previous referential NP). Furthermore, different referential devices are sensitive to different combinations of these factors. This conclusion extends the domain of a single language to other languages, such that the set of relevant factors for a certain referential device can differ depending on a language. Taken together, this form-specific multiple constraints approach, which also takes the language’s individual features into account is deemed to be pertinent for explaining a language’s system of referential devices.

More specifically, the present analysis focuses on four aspects of referential expressions: i) zero reference and personal pronouns as devices of minimal reference, ii) article-like determiners as a means of expressing definiteness and



indefiniteness in non-article languages, iii) pronouns and their division of labour in discourse, and iv) the diverging functions of demonstrative adverbs and demonstrative pronouns as determiners. The main conclusions of these individual studies can be summarised as follows.

**First**, as discussed in ZERO, zero reference has more constraints than the third person overt personal pronoun *ta* in Estonian, although these forms often occur in linguistically rather similar environments. That is, both devices refer to already mentioned and salient entities in discourse and usually signal the CONTINUE transition in terms of Centering theory (Grosz, Joshi, and Weinstein 1995), meaning that they have similar effect on the overall coherence of discourse. What is intrinsic to zero reference is that it clearly prefers nominal subjects and main clauses and, according to Centering theory, should require very little cognitive effort from speech participants during the production as well as processing stages. These restrictions do not apply to the overt personal pronoun, meaning that the overt form is plausible in all contexts in which zero reference fails, and furthermore, *ta* can often be substituted for zero reference without consequences on the discourse's coherence. Therefore, *ta* 's/he' is considered to be the most neutral, default anaphoric device in Estonian when referring to most salient discourse entities. The use of zero reference, on the other hand, depends on the speaker's individual preferences.

**Second**, the data presented in DETERMINERS corroborate the previous findings that in Estonian, article-like determiners are not as frequent as noun phrases without determiners and that determiners are not grammatically mandatory in any context. However, the two most frequent determiners, indefinite *üks* 'one/a' and definite *see* 'this/the', exhibit the potential to develop into more grammatical devices such as articles. This belief is supported by the higher frequency of determiners in Estonian data as compared to Finnish and Russian data (two other article-less languages). Also, the determiners *üks* and *see* express a clear-cut work allocation that directly connects to definiteness-marking: the determiner *üks* is restricted to first mentions (i.e., indefiniteness), while *see* coincides with subsequent mentions (i.e., definiteness). The indefinite determiner *üks* exhibits an asymmetry regarding animacy, a semantic feature, so that the determiner is more characteristic to animate as compared to inanimate reference. This indicates that at the current stage, *üks* 'one/a' is still more sensitive to semantic selection criteria than grammatical ones. The definite determiner *see*, on the other hand, is not restricted to certain linguistic or discourse environments. This circumstance signals that the definite determiner *see* can occur across different contexts of definiteness marking paradigm. Taken together, this thesis supports the view that *üks* and *see* are currently at the initial stages of grammaticalization process (Heine and Kuteva 2006).

**Third**, the study PRONOUNS found that anaphoric pronouns are widely used in Estonian spoken narratives, but the more specific division into pronoun types shows that personal pronouns and zero reference are used substantially more often (90% of anaphoric pronouns) than demonstrative pronouns (8.5%), whereas anaphoric demonstrative adverbs are rare (1.6%) and restricted to referring to

spatial relations in text. The results indicate that in Estonian, the third person pronoun *ta* and zero reference stand much closer to each other than the personal pronoun and demonstrative pronoun. For example, both *ta* and zero reference are common in referring to animate entities when referential distance is small. The demonstrative pronoun, on the other hand, mostly refers to inanimate entities, although it can be used for animate reference when there are multiple animate entities in a clause. Importantly, the factors that determine the pronoun choice are different across pronoun types, indicating that the distribution of pronoun types obtained in this data is not absolute and may be somewhat different in a context in which the influential factors have different emphasis.

And **fourth**, it is shown in DEMONSTRATIVES that meaning-related factors help to explain certain aspects of the variation in determiner forms. As was already discussed in DETERMINERS, Estonian speakers use certain pronouns for article-like functions. An interesting feature of Estonian is that, in certain contexts, demonstrative adverbs (e.g. *siin* ‘here’, *seal* ‘there’) can occupy the determiner position in a noun phrase. The study shows that the semantic class of the head noun and noun concreteness can predict whether a demonstrative pronoun or a demonstrative adverb is chosen as a (definite) determiner in a particular context. Nouns expressing a spatial meaning prefer demonstrative adverbs as determiners, while non-spatial meaning combines with demonstrative pronouns. The effect of noun concreteness is not so straightforward, but the general tendency appears to be that demonstrative adverbs as determiners are more common with concrete nouns. These outcomes are especially valuable for interpreting the usage of polysemous nouns since they allow one to explain when a noun should be understood as a place (with an adverbial determiner) and when should it be conceptualized as an entity (with a pronominal determiner). Isolated nouns themselves, without a context, have only potential referents. This potential gets realized in actual communication, when a certain linguistic (or non-linguistic) device is used (Taremaa et al. 2021). Taken together, it is concluded that the choice of a determiner form contributes to language’s referential potentiality.

In addition to Estonian-centred results, the thesis opens up a number of **cross-linguistic implications**. Essentially, overall usage frequencies and contexts of different types of referential devices (e.g., full NPs with determiners, pronominals) along with the set of factors that determine the occurrence of this device in discourse reveal that in addition to varying within a language, particular devices may also have distinct functions across languages. Thus, it follows that even when devices carry the same descriptive label in grammars (e.g., demonstrative pronoun), this does not suggest that these devices also have similar referential properties across languages. Instead, as this thesis has demonstrated, languages exhibit diverse referential practices. For example, the results support the account that languages can be divided along the continuum of demonstrative use, where some languages are demonstrative oriented (e.g., Finnish) and some do not rely on demonstratives (e.g., Russian) (Pajusalu et al. 2018). Also, even in languages that do not possess grammatical articles, the expression of identifiability with

pragmatic means seems to be a common necessity. However, languages differ in terms of how frequently and in which contexts the identifiability of a referent is also formally marked, and which devices function as determiners, i.e., markers of identifiability.

In sum, this thesis provides a systematic and multifactorial insight into the system of Estonian anaphoric referential devices. In addition to looking at the behaviour of individual devices in their natural usage contexts, an overall picture is outlined about how all these devices work together in discourse. Furthermore, the claims made in this thesis go beyond Estonian and suggest some methodological improvements for studying reference in other languages as well. The findings of the thesis offer substantial knowledge about how different linguistic expressions can be used for referring to entities around us and which underlying factors affect the choice of a particular expression, and how. It is hoped that the outcome brings us a step closer to understanding how human language and its discourse processing mechanisms help us to perform in this complicated world.

## REFERENCES

- Abbott, Barbara. 2006. Definite and indefinite. In Keith Brown (ed.), *Encyclopedia of Language & Linguistics (Second Edition)*, 392–399. Elsevier Science.  
<https://doi.org/10.1016/B0-08-044854-2/01089-0>.
- Abbott, Barbara. 2010. *Reference*. Oxford: Oxford University Press.
- Amon, Marri. 2015. *Initial and final detachments in spoken Estonian: a study in the framework of Information Structuring*. (Dissertationes Linguisticae Universitatis Tartuensis 24). Tartu: University of Tartu Press.
- Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. London: Routledge.
- Ariel, Mira. 1994. Interpreting anaphoric expressions: A cognitive versus a pragmatic approach. *Journal of Linguistics* 30(1). 3–42. <https://doi.org/10.2307/4176255>.
- Ariel, Mira. 2001. Accessibility theory: An overview. In Ted J. M. Sanders, Joost Schilperoord and Wilbert Spooren (eds.), *Text representation: Linguistic and psycholinguistic aspects*, 29–87. Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://doi.org/https://doi.org/10.1075/hcp.8.04ari>.
- Arnold, Jennifer E. 2001. The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes* 31(2). 137–162.  
[https://doi.org/10.1207/S15326950DP3102\\_02](https://doi.org/10.1207/S15326950DP3102_02).
- Arnold, Jennifer E. 2003. Multiple constraints on reference form: Null, pronominal, and full reference in Mapudungun. In John W. Du Bois, Lorraine E. Kumpf and William J. Ashby (eds.), *Preferred Argument Structure: Grammar as architecture for function*, 225–246. Amsterdam/Philadelphia: John Benjamins.
- Arnold, Jennifer E. 2010. *How speakers refer: The role of accessibility*. *Language and Linguistics Compass* 4. 187–203.  
<https://doi.org/10.1111/j.1749-818X.2010.00193.x>.
- Arnold, Jennifer E. and Zenzi M. Griffin. 2007. The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language* 56. 521–536. <https://doi.org/10.1016/j.jml.2006.09.007>.
- Arnold, Jennifer E. and Sandra A. Zerkle. 2019. Why do people produce pronouns? Pragmatic selection vs. rational models. *Language, Cognition and Neuroscience* 34(9). 1152–1175. <https://doi.org/10.1080/23273798.2019.1636103>.
- Arppe, Antti. 2009. Linguistic choices vs. probabilities – how much and what can linguistic theory explain? In Sam Featherston and Susanne Winkler (eds.), *Volume 1 Process*, 1–24. Berlin/New York: De Gruyter Mouton.  
<https://doi.org/10.1515/9783110216141.1>.
- Baayen, R. Harald, Anna Endresen, Laura A. Janda, Anastasia Makarova and Tore Nesset. 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>.
- Berman, Ruth and Dan Slobin (eds.). 1994. *Relating events in narrative: A cross-linguistic developmental study*. Hillsdale, NJ: Erlbaum.
- Binder, Jeffrey R., Lisa L. Conant, Colin J. Humphries, Leonardo Fernandino, Stephen B. Simons, Mario Aguilar and Rutvik H. Desai. 2016. Toward a brain-based componential semantic representation. *Cognitive Neuropsychology* 33(3–4). 130–174. <https://doi.org/10.1080/02643294.2016.1147426>.
- Boas, Hans C. 2008. Determining the structure of lexical entries and grammatical constructions in Construction Grammar. In *Annual review of cognitive linguistics* 6, 113–144.

- Breiman, Leo. 2001. Random forests. *Machine Learning* 45(1). 5–32.  
<https://doi.org/10.1023/A:1010933404324>.
- Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Sam Featherston and Wolfgang Sternfeld (eds.), *Roots: Linguistics in search of its evidential base*, 77–96. Berlin: Mouton De Gruyter.
- Brown-Schmidt, Sarah, Donna K. Byron and Michael K. Tanenhaus. 2005. Beyond salience: Interpretation of personal and demonstrative pronouns. *Journal of Memory and Language* 53. 292–313. <https://doi.org/10.1016/j.jml.2005.03.003>.
- Bühler, Karl. 1990 [1934]. *Theory of language: the representational function of language*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and topic*, 25–55. New York: Academic Press.
- Chafe, Wallace L. (ed.). 1980. *The pear stories: cognitive, cultural, and linguistic aspects of narrative production*. Norwood, N.J: Ablex.
- Chafe, Wallace L. 1994. *Discourse, consciousness, and time: the flow and displacement of conscious experience in speaking and writing*. Chicago/London: The University of Chicago Press.
- Chambers, Craig G. and Ron Smyth. 1998. Structural parallelism and discourse coherence: A test of Centering theory. *Journal of Memory and Language* 39(4). 593–608.
- Chesterman, Andrew. 1991. *On definiteness. A study with special reference to English and Finnish*. Cambridge: Cambridge University Press.
- Colonna, Saveria, Sarah Schimke and Barbara Hemforth. 2012. Information structure effects on anaphora resolution in German and French: A crosslinguistic study of pronoun resolution. *Linguistics* 50(5). 991–1013.  
<https://doi.org/10.1515/ling-2012-0031>.
- Conover, J.W. 1999. *Practical nonparametric statistics (3rd ed.)*. New York: Wiley.
- Cornish, Francis. 1999. *Anaphora, discourse, and understanding: Evidence from English and French*. Oxford: Clarendon Press.
- Coventry, Kenny R., Debra Griffiths and Colin J. Hamilton. 2014. Spatial demonstratives and perceptual space: Describing and remembering object location. *Cognitive Psychology* 69. 46–70. <https://doi.org/10.1016/j.cogpsych.2013.12.001>.
- Coventry, Kenny R., Berenice Valdés, Alejandro Castillo and Pedro Guijarro-Fuentes. 2008. Language within your reach: Near-far perceptual space and spatial demonstratives. *Cognition* 108(3). 889–895. <https://doi.org/10.1016/j.cognition.2008.06.010>.
- Crawley, Rosalind A. 1986. Some factors influencing the comprehension of pronouns. In Charles Clifton (ed.), *Proceedings of the eighth annual conference of the Cognitive Science Society*, 613–620. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Crawley, Rosalind A., Rosemary J. Stevenson and David Kleinman. 1990. The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research* 19(4). 245–264.
- Croft, William. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Crystal, David. 2008. *A dictionary of linguistics and phonetics*. Oxford, UK: Blackwell Publishing Ltd. <https://doi.org/10.1002/9781444302776>.
- Dancygier, Barbara. 2019. Proximal and distal deictics and the construal of narrative time. *Cognitive Linguistics* 30(2). 399–415. <https://doi.org/10.1515/cog-2018-0044>.

- Diessel, Holger. 1999. *Demonstratives: Form, function, and grammaticalization*. (Typological Studies in Language 42.) Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Diessel, Holger. 2006. Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics* 17(4). 463–489. <https://doi.org/10.1515/COG.2006.015>.
- Diessel, Holger. 2013. Distance contrasts in demonstratives. In Matthew S. Dryer and Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Di Eugenio, Barbara. 1998. Centering in Italian. In Marilyn A. Walker, Aravind K. Joshi and Ellen F Prince (eds.), *Centering theory in discourse*, 115–137. Oxford: Clarendon Press.
- Disbray, Samantha. 2016. The development of reference realization and narrative in an Australian contact language, Wumpurrarni English. *Frontiers in Psychology* 7(43). <https://doi.org/10.3389/fpsyg.2016.00043>.
- Dryer, Matthew S. 2013a. Indefinite articles. In Matthew S. Dryer and Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/38>.
- Dryer, Matthew S. 2013b. Definite articles. In Matthew S. Dryer and Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/37>.
- Dryer, Matthew S. 2013c. Order of subject, object, and verb. In Matthew S. Dryer and Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/81>.
- Du Bois, John W. 1980a. Beyond definiteness: The trace of identity in discourse. In Wallace L. Chafe (ed.), *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*, 203–274. Norwood, N.J: Ablex.
- Du Bois, John W. 1980b. The search for a cultural niche: Showing the Pear Film in a Mayan Community. In Wallace L. Chafe (ed.), *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*, 1–7. Norwood, NJ: Ablex.
- Du Bois, John W. 1987. The discourse basis of ergativity. *Language* 63(4). 805–855.
- Du Bois, John W. 2003. Argument structure: grammar in use. In John W. Du Bois, Lorraine E. Kumpf and William J. Asby (eds.), *Preferred Argument Structure: Grammar as architecture for function*, 11–60. (Studies in Discourse and Grammar 14). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Duguine, Maia. 2014. Argument ellipsis: A unitary approach to pro-drop. *Linguistic Review* 31(3). 515–549. <https://doi.org/10.1515/tlr-2014-0010>.
- Dumont, Jenny. 2016. *Third person references. Forms and functions in two spoken genres of Spanish*. (Studies in Functional and Structural Linguistics 71.) Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/sfsl.71>.
- Duvallon, Outi and Antoine Chalvin. 2004. La réalisation zéro du pronom sujet de première et de deuxième personne du singulier en finnois et en estonien parlés. *Linguistica Uralica* XL(4). 270–286.
- Erelt, Mati. 2009. Typological overview of Estonian syntax. *STUF - Language Typology and Universals* 62(1/2). 6–28. <https://doi.org/10.1524/stuf.2009.0002>.
- Erelt, Mati. 2017a. Sissejuhatus süntaksisse [Introduction to syntax]. In Mati Erelt and Helle Metslang (eds.), *Eesti keele süntaks [The syntax of Estonian]*. (*Eesti keele varamu III*), 53–89. Tartu: University of Tartu Press.

- Erelt, Mati. 2017b. Ellipsis [Ellipsis]. In Mati Erelt and Helle Metslang (eds.), *Eesti keele süntaks [The syntax of Estonian], (Eesti keele varamu III)*, 590–602. Tartu: University of Tartu Press.
- Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael and Silvi Vare. 1993. *Eesti keele grammatika II. Süntaks, lisa: kiri [Estonian grammar II. Syntax. Appendix: Orthography]*. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.
- Erelt, Mati and Helle Metslang. 2006. Estonian clause patterns - From Finno-Ugric to Standard Average European. *Linguistica Uralica* XLII(4). 254–266.
- Etelämäki, Marja. 2009. The Finnish demonstrative pronouns in light of interaction. *Journal of Pragmatics* 41(1). 25–46. <https://doi.org/10.1016/j.pragma.2008.09.005>.
- Faure, Richard. 2014. Determiners. In Giorgos K. Giannakis (ed.), *Encyclopedia of Ancient Greek Language and Linguistics*. Brill. [http://dx.doi.org/10.1163/2214-448X\\_eagll\\_COM\\_00000094](http://dx.doi.org/10.1163/2214-448X_eagll_COM_00000094).
- Fedele, Emily and Elsi Kaiser. 2014. Looking back and looking forward: Anaphora and cataphora in Italian. *University of Pennsylvania Working Papers in Linguistics* 20(1). 81–90.
- Fedorova, Olga V. 2014. The role of potential referential conflict in the choice of a referring expression. *The Russian Journal of Cognitive Science* 1(1–2). 6–21.
- Filiaci, Francesca, Antonella Sorace and Manuel Carreiras. 2014. Anaphoric biases of null and overt subjects in Italian and Spanish: a cross-linguistic comparison. *Language, Cognition and Neuroscience*. Taylor & Francis 29(7). 825–843. <https://doi.org/10.1080/01690965.2013.801502>.
- Fillmore, Charles J. 1997. *Lectures on deixis*. Stanford: CSLI Publications.
- Fraurud, Kari. 1990. Definiteness and the processing of NP's in natural discourse. *Journal of Semantics* 7(4). 395–433. <http://dx.doi.org/10.1093/jos/7.4.395>.
- Freienthal, Linda. 2020. *Pronominaalsete viitesuhete automaatne lahendamise eesti keeles närvivõrkude abil [Pronominal coreference resolution in Estonian with neural networks]*. (Master's thesis). University of Tartu.
- Fukumura, Kumiko and Roger van Gompel. 2015. Effects of order of mention and grammatical role on anaphor resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 41(2). 501–525. <https://doi.org/10.1037/xlm0000041>.
- Garnham, Alan. 2001. *Mental models and the interpretation of anaphora. Essays in cognitive psychology*. Philadelphia: Psychology Press.
- Gelormini-Lezama, Carlos and Amit Almor. 2011. Repeated names, overt pronouns, and null pronouns in Spanish. *Language and Cognitive Processes* 26(3). 437–454. <https://doi.org/10.1080/01690965.2010.495234>.
- Gernsbacher, Morton Ann. 1990. *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gilquin, Gaëtanelle and Stefan Th. Gries. 2009. Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory* 5(1). 1–26. <https://doi.org/10.1515/CLLT.2009.001>.
- Givón, Talmy. 1981. On the development of the numeral “one” as an indefinite marker. *Folia Linguistica Historica* 2(1). 35–53.
- Givón, Talmy. 1983a. Topic continuity in discourse: An introduction. In Talmy Givón (ed.), *Topic continuity in discourse: A quantitative cross-language study*, 1–42. Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Givón, Talmy. 1983b. Topic continuity in spoken English. In Talmy Givón (ed.), *Topic continuity in discourse: A quantitative cross-language study*, 345–363. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Givón, Talmy. 1993. Coherence in text, coherence in mind. *Pragmatics and Cognition* 1(2). 171–227.
- Givón, Talmy. 1995. *Functionalism and grammar*. Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://doi.org/https://doi.org/10.1075/z.74>.
- Givón, Talmy. 2005. *Context as Other Minds. The Pragmatics of Sociality, Cognition and Communication*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago/London: University of Chicago Press.
- Goldberg, Adele E (ed.). 1996. *Conceptual structure, discourse and language*. Stanford: Center for the Study of Language and Information.
- Goldberg, Adele E. 2006. *Constructions at work: The nature of generalizations in language*. Oxford: Oxford University Press.
- Gonzalez-Marquez, Monica, Irene Mittelberg, Seana Coulson and Michael J. Spivey (eds.). 2007. *Methods in Cognitive Linguistics*. (Human Cognitive Processing 18). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Gordon, Peter C., Bj Grosz and La Gilliom. 1993. Pronouns, names, and the centering of attention in discourse. *Cognitive science* 17. 311–347. [https://doi.org/10.1207/s15516709cog1703\\_1](https://doi.org/10.1207/s15516709cog1703_1).
- Grafmiller, Jason. 2014. Variation in English genitives across modality and genres. *English Language and Linguistics* 18(3). 471–496. <https://doi.org/10.1017/S1360674314000136>.
- Greenberg, Joseph H. 1978. How does a language acquire gender markers? In Joseph H. Greenberg (ed.), *Universals of Human Language*, vol. III, 47–82. Stanford: Stanford University Press.
- Greene, S.B., R.J. Gerrig, G. Mckoon and R. Ratcliff. 1994. Unheralded pronouns and management by common ground. *Journal of Memory and Language* 33(4). 511–526. <https://doi.org/10.1006/jmla.1994.1024>.
- Gries, Stefan Th. 2001. A multifactorial analysis of syntactic variation: Particle movement revisited. *Journal of Quantitative Linguistics* 8(1). 33–50. <https://doi.org/10.1076/jqul.8.1.33.4092>.
- Gries, Stefan Th. 2003. *Multifactorial analysis in corpus linguistics: A study of particle placement*. London/New York: Continuum Press.
- Gries, Stefan Th. 2017. Syntactic alternation research: Taking stock and some suggestions for the future. *Belgian Journal of Linguistics* 31(2). 8–29. <https://doi.org/10.1075/bjl.00001.gri>.
- Gries, Stefan Th. and Anatol Stefanowitsch. 2004. Extending collocation analysis. *International Journal of Corpus Linguistics* 9(1). 97–129. <https://doi.org/10.1075/ijcl.9.1.06gri>.
- Gries, Stefan Th. and Nick C. Ellis. 2015. Statistical measures for usage-based linguistics. *Language Learning* 65(S1). 228–255. <https://doi.org/10.1111/lang.12119>.
- Grosz, Barbara J., Aravind K. Joshi and Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21(2). 203–225.



- Gundel, Jeanette K. 1998. Centering theory and the Givenness Hierarchy: Towards a synthesis. In Marilyn A. Walker, Aravind K. Joshi and Ellen F. Prince (eds.), *Centering theory in discourse*, 183–198. Oxford: Oxford University Press.
- Gundel, Jeanette K. and Barbara Abbott. 2019. Introduction. In Jeanette K. Gundel and Barbara Abbott (eds.), *The Oxford Handbook of Reference*. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199687305.013.1>.
- Gundel, Jeanette K., Mamadou Bassene, Bryan Gordon, Linda Humnick and Amel Khalfaoui. 2010. Testing predictions of the Givenness Hierarchy framework: A crosslinguistic investigation. *Journal of Pragmatics* 42(7). 1770–1785. <https://doi.org/10.1016/j.pragma.2009.09.010>.
- Gundel, Jeanette K., Nancy Hedberg and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307. <https://doi.org/10.2307/416535>.
- Gundel, Jeanette K., Nancy Hedberg and Ron Zacharski. 2012. Underspecification of cognitive status in reference production: some empirical predictions. *Topics in cognitive science* 4(2). 249–68. <https://doi.org/10.1111/j.1756-8765.2012.01184.x>.
- Gundel, Jeanette K. and Kaitlin Johnson. 2013. Children’s use of referring expressions in spontaneous discourse: Implications for theory of mind development. *Journal of Pragmatics* 56. 43–57. <https://doi.org/10.1016/j.pragma.2013.04.003>.
- Haegeman, Liliane. 2007. Subject omission in present-day written English – on the theoretical relevance of peripheral data. *Revista di Grammatica Generativa* 32. 91–124.
- Haegeman, Liliane. 2013. The syntax of registers: Diary subject omission and the privilege of the root. *Lingua* 130. 88–110. <https://doi.org/10.1016/j.lingua.2013.01.005>.
- Hakulinen, Auli. 1985. On cohesive devices in Finnish. In E. Sözer (ed.), *Text connectivity, text coherence. Aspects, methods, results*, (Papers in textlinguistics 49), 337–384. Hamburg: Helmut Buske Verlag.
- Halliday, Michael A. K. 1985. *An introduction to functional grammar. Second Edition*. London: Edward Arnold.
- Halliday, Michael A. K. and Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
- Hanks, Peter. 2019. Reference as a speech act. In Jeanette K. Gundel and Barbara Abbott (eds.), *The Oxford Handbook of Reference*. Oxford, UK: Oxford University Press.
- Hanks, William F. 1990. *Referential practice: Language and lived space among the Maya*. Chicago/London: University of Chicago Press.
- Hanks, William F. 1992. The indexical ground of deictic reference. In Alessandro Duranti and Charles Goodwin (eds.), *Rethinking context: Language as an interactive phenomenon*. Cambridge: Cambridge University Press.
- Haspelmath, Martin. 1998. How young is standard average european? *Language Sciences* 20(3). 271–287. [https://doi.org/10.1016/S0388-0001\(98\)00004-7](https://doi.org/10.1016/S0388-0001(98)00004-7).
- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687. <https://doi.org/10.1353/lan.2010.0021>.
- Hedberg, Nancy. 2010. Centering and noun phrase realization in Kaqchikel Mayan. *Journal of Pragmatics* 42(7). 1829–1841. <https://doi.org/10.1016/j.pragma.2009.09.011>.
- Heine, Bernd, UlrikeClaudi and Friederike Hünemeyer. 1991. *Grammaticalization: A conceptual framework*. Chicago: University of Chicago Press.

- Heine, Bernd and Tania Kuteva. 2006. *The changing languages of Europe*. New York: Oxford University Press.
- Helasvuo, Marja-Liisa. 1993. Are relative clauses either restrictive or non-restrictive? A study of the relative clauses in the Finnish Pear Stories. In Anders Holmberg and Urpo Nikanne (eds.), *Case and other functional categories in Finnish syntax* (Studies in Generative Grammar 39), 163–176. Berlin/New York: De Gruyter Mouton. <https://doi.org/https://doi.org/10.1515/9783110902600.163>.
- Helasvuo, Marja-Liisa. 2001. *Syntax in the Making*. (Studies in Discourse and Grammar 9). Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/sidag.9>.
- Helasvuo, Marja-Liisa. 2003. Argument splits in Finnish grammar and discourse. In John W. Du Bois, Lorraine E. Kumpf and William J. Ashby (eds.), *Preferred Argument Structure: Grammar as architecture for function*, 247–272. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Helasvuo, Marja-Liisa. 2020. The noun phrase as an emergent unit in Finnish. In Tsuyoshi Ono and Sandra A. Thompson (eds.), *The 'Noun Phrase' across languages. An emergent unit in interaction* (Typological Studies in Language 128), 71–92. Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/tsl.128.04hel>.
- Hendriks, Petra. 2016. Cognitive modeling of individual variation in reference production and comprehension. *Frontiers in Psychology* 7. 506. <https://doi.org/10.3389/fpsyg.2016.00506>.
- Hennoste, Tiit. 1986. Üldsõnad ja kohesioon [General words and cohesion]. In *Sõnasõel: uurimusi ja materjale eesti keele sõnavara alalt* 7, 46–58. University of Tartu.
- Hickmann, Maya. 2004. Coherence, cohesion, and context: some comparative perspectives in narrative development. In S. Strömquist and L. Verhoeven (eds.), *Relating events in narrative: Typological and contextual perspectives*, 281–306. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Hiietam, Katrin. 2003. *Definiteness and grammatical relations in Estonian*. (Doctoral thesis.) University of Manchester.
- Himmelmann, Nikolaus P. 1996. Demonstratives in narrative discourse: a taxonomy of universal uses. In Barbara Fox (ed.), *Studies in Anaphora*, 205–254. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Hint, Helen and Elsi Kaiser. 2021. Overt and zero third person reference in Estonian: When and why. Manuscript in preparation.
- Hint, Helen, Maria Reile and Renate Pajusalu. 2013. Kontekst ja viitamine: argivestlused, legod ja narratiivid [Context and reference: everyday conversations, Lego blocks and narratives]. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics* 3. 161–183. <https://doi.org/10.12697/jeful.2013.4.1.10>.
- Hobbs, Jerry R. 1979. Coherence and coreference. *Cognitive Science* 3. 67–90.
- Hoek, Jet, Hannah Rohde, Jacqueline Evers-Vermeul and Ted J. M. Sanders. 2020. Scolding the child who threw the scissors: Shaping discourse expectations by restricting referents. *Language, Cognition and Neuroscience* 36(3). 382–399. <https://doi.org/10.1080/23273798.2020.1852292>.
- Hoek, Karen van. 1997. *Anaphora and Conceptual Structure*. Chicago and London: The University of Chicago Press.

- Hoffmann, Beryl. 1998. Word order, information structure, and Centering in Turkish. In Marilyn A. Walker, Aravind K. Joshi and Ellen F Prince (eds.), *Centering theory in discourse*, 253–271. Oxford: Clarendon Press.
- Hothorn, Torsten, Kurt Hornik, Carolin Strobl and Achim Zeileis. 2020. party: A laboratory for recursive partytioning. <https://cran.r-project.org/package=party>.
- Hothorn, Torsten, Kurt Hornik and 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>.
- Huang, Yan. 2000. Discourse anaphora: Four theoretical models. *Journal of Pragmatics* 32(2). 151–176. [https://doi.org/10.1016/S0378-2166\(99\)00041-7](https://doi.org/10.1016/S0378-2166(99)00041-7).
- Hurewitz, Felicia. 1998. A quantitative look at discourse coherence. In Marilyn Walker, Aravind K. Joshi and Ellen F Prince (eds.), *Centering Theory in discourse*, 273–291. Oxford: Clarendon Press.
- Jackendoff, Ray. 2002. *Foundations of language: Brain, meaning, grammar, evolution*. Oxford/New York: Oxford University Press.  
<https://doi.org/10.1093/acprof:oso/9780198270126.001.0001>.
- Janda, Laura A. 2013. Quantitative methods in cognitive linguistics: An introduction. In Laura A. Janda (ed.), *Cognitive linguistics: The quantitative turn. The essential reader*, 1–32. Berlin/Boston: De Gruyter Mouton.
- Janda, Laura A. and Valery D. Solovyev. 2009. What constructional profiles reveal about synonymy: A case study of Russian words for SADNESS and HAPPINESS. 20(2). 367–393. <https://doi.org/10.1515/9783110335255.295>.
- Jarbou, Samir Omar. 2010. Accessibility vs. physical proximity: An analysis of exophoric demonstrative practice in Spoken Jordanian Arabic. *Journal of Pragmatics* 42(11). 3078–3097. <https://doi.org/10.1016/j.pragma.2010.04.014>.
- Järvikivi, Juhani, Roger P. G. van Gompel, Jukka Hyönä and Raymond Bertram. 2005. Ambiguous pronoun resolution. Contrasting the first-mention and subject-preference accounts. *Psychological Science* 16(4). 260–264. <https://doi.org/10.1111/j.0956-7976.2005.01525.x>.
- 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. 2011a. Focusing on pronouns: Consequences of subjecthood, pronominalisation, and contrastive focus. *Language and Cognitive Processes* 26(10). 1625–1666. <https://doi.org/10.1080/01690965.2010.523082>.
- Kaiser, Elsi. 2011b. Salience and contrast effects in reference resolution: The interpretation of Dutch pronouns and demonstratives. *Language and Cognitive Processes* 26(10). 1587–1624. <https://doi.org/10.1080/01690965.2010.522915>.
- Kaiser, Elsi. 2013. Looking beyond personal pronouns and beyond English: Typological and computational complexity in reference resolution. *Theoretical Linguistics* 39(1–2). 109–122. <https://doi.org/10.1515/tl-2013-0007>.
- Kaiser, Elsi. 2015. Impersonal and generic reference: A cross-linguistic look at Finnish and English narratives. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics* 6(2). 9–42.
- Kaiser, Elsi and Katrin Hiietam. 2003. A comparison of the referential properties of third person pronouns in Finnish and Estonian. *Nordlyd. Proceedings of the Workshop on Generative Approaches to Finnic Languages* 31(4). 654–667.

- Kaiser, Elsi and John C. Trueswell. 2008. Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes* 23(5). 709–748.  
<https://doi.org/10.1080/01690960701771220>.
- Kaiser, Elsi and Virve-Anneli Vihman. 2006. Invisible arguments. Effects of demotion in Estonian and Finnish. In B. Lyngfelt and T. Solstad (eds.), *Demoting the Agent: Passive, middle and other voice phenomena*, 111–141. Amsterdam: John Benjamins.
- Kaiser, Elsi and 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.
- Karimi, Hossein, Kumiko Fukumura, Fernanda Ferreira and Martin J. Pickering. 2014. The effect of noun phrase length on the form of referring expressions. *Memory and Cognition* 42(6). 993–1009. <https://doi.org/10.3758/s13421-014-0400-7>.
- Kärkkäinen, Elise. 1996. Preferred argument structure and subject role in American English conversational discourse. *Journal of Pragmatics* 25(5). 675–701.  
[https://doi.org/10.1016/0378-2166\(95\)00010-0](https://doi.org/10.1016/0378-2166(95)00010-0).
- Keevallik, Leelo. 2003. *From interaction to grammar: Estonian finite verb forms in conversation*. Uppsala: Uppsala University.
- Keevallik, Leelo. 2010. The interactional profile of a placeholder: the Estonian demonstrative *see*. In Nino Amiridze, Boyd H. Davis and Margaret MacLagan (eds.), *Fillers, pauses, and placeholders*, 139–172. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Kehler, Andrew, Laura Kertz, Hannah Rohde and Jeffrey L. Elman. 2008. Coherence and coreference revisited. *Journal of Semantics* 25. 1–44.  
<https://doi.org/10.1093/jos/ffm018>.
- Kehler, Andrew and Hannah Rohde. 2013. A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics* 39. 1–37. <https://doi.org/10.1515/tl-2013-0001>.
- Kibrik, Andrej A. 1996. Anaphora in Russian narrative prose: A cognitive calculative account. In Barbara Fox (ed.), *Studies in anaphora*, 255–303. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Kibrik, Andrej A. 2011. *Reference in discourse. (Oxford Studies in Typology and Linguistic Theory)*. Oxford/New York: Oxford University Press.
- Kibrik, Andrej A., Mariya V. Khudyakova, Grigory B. Dobrov, Anastasia Linnik and Dmitrij A. Zalmanov. 2016. Referential choice: Predictability and its limits. *Frontiers in Psychology* 7. 1429. <https://doi.org/10.3389/fpsyg.2016.01429>.
- Kibrik, Andrej A. 2019. Rethinking agreement: Cognition-to-form mapping. *Cognitive Linguistics* 30(1). 37–83. <https://doi.org/10.1515/cog-2017-0035>.
- Kirsipuu, Helen. 2012. *Referentsiaalsete noomenifraaside omadused kirjalikes narratiivides [Referential properties of noun phrases in Estonian written narratives]*. (Master’s thesis). University of Tartu.
- Kirsipuu, Helen, Piret Soodla and Renate Pajusalu. 2012. Referentsiaalsed noomenifraasid laste narratiivides [Referential noun phrases in Estonian children’s narratives]. *Eesti Rakenduslingvistika Ühingu aastaraamat = Estonian papers in applied linguistics* 8. 91–107. <https://doi.org/10.5128/ERYa8.06>.
- Kivik, Piibi-Kai. 2010. Personal pronoun variation in language contact: Estonian in the United States. In Muriel Norde, Bob de Jonge and Cornelius Hasselblatt (eds.), *Language Contact: New Perspectives*, 63–86. Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Klavan, Jane. 2012. *Evidence in linguistics: corpus-linguistic and experimental methods for studying grammatical synonymy*. (Dissertationes Linguisticae Universitatis Tartuens 15). Tartu: University of Tartu Press.
- Koster, Charlotte, John Hoeks and Petra Hendriks. 2011. Comprehension and production of subject pronouns in child Dutch. In Angela Grimm, Anja Müller, Cornelia Hamann and Esther Ruigendijk (eds.), *Production-comprehension asymmetries in child language*, 99–122. Berlin/Boston: De Gruyter Mouton.
- Krasavina, Olga. 2011. Demonstratives and salience: Towards a functional taxonomy. In Christian Chiarcos, Berry Claus and Michael Grabski (eds.), *Salience: Multi-disciplinary perspectives on its function in discourse*, 31–55. Berlin/New York: De Gruyter Mouton.
- Kwon, Nayoung and Patrick Sturt. 2013. Null pronominal (pro) resolution in Korean, a discourse-oriented language. *Language and Cognitive Processes* 28(3). 377–387. <https://doi.org/10.1080/01690965.2011.645314>.
- Labov, William. 1972. *Language in the inner city: Studies in the Black English vernacular*. Philadelphia: University of Pennsylvania Press.
- Labov, William and Joshua Waletzky. 1967. Narrative analysis: Oral versions of personal experience. In June Helm (ed.), *Essays on the verbal and visual arts*, 12–44. Seattle/London: University of Washington Press.
- Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Lakoff, George. 1990. The invariance hypothesis: is abstract reason based on image-schemas? *Cognitive Linguistics* 1(1). 39–74. <https://doi.org/10.1515/cogl.1990.1.1.39>.
- Lambrecht, Knud. 1994. *Information structure and sentence form. Topic, focus and the mental representation of discourse referents*. Cambridge: Cambridge University Press.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Volume I: Theoretical prerequisites*. Stanford: Stanford University Press.
- Laury, Ritva. 1991. On the development of the definite article *se* in spoken Finnish. *SKY – Yearbook of the Linguistic Association of Finland* 4. 93–121.
- Laury, Ritva. 1994. *Se* ensinmaininnan pronominin puhutussa suomessa. *Virittäjä* 98(3). 449–457.
- Laury, Ritva. 1996a. Pronouns and adverbs, figure and ground: The local case forms and locative forms of the Finnish demonstratives in spoken discourse. *SKY – Yearbook of the Linguistic Association of Finland* 9. 65–92.
- Laury, Ritva. 1996b. *Sen* kategoriasta – onko suomessa jo artikkeli? *Virittäjä* 100(2). 162–181.
- Laury, Ritva. 1997. *Demonstratives in interaction: the emergence of a definite article in Finnish*. (Studies in Discourse and Grammar 7). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Laury, Ritva. 2001. Definiteness. In Jef Verschueren, Jan-Ola Östman, Jan Blommaert and Chris Bulcaen (eds.), *Handbook of pragmatics. Volume 5*. Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://doi.org/https://doi.org/10.1075/hop.5.def1>.
- Laury, Ritva. 2005. Introduction. In Ritva Laury (ed.), *Minimal reference: The use of pronouns in Finnish and Estonian discourse*. (Studia Fennica Linguistica 12.), 7–10. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Lehmann, Christian. 2004. Data in linguistics. *The Linguistic Review* 21(3/4). 275–310. [10.1515/tlir.2004.21.3-4.175](https://doi.org/10.1515/tlir.2004.21.3-4.175).

- Lepajõe, Kersti. 2004. Isikudeiksis õpilastekstides [Person deixis in students' texts]. In Reet Kasik (ed.), *Tekstid ja taustad III. Lingvistiline tekstianalüüs*, 74–89. Tartu: University of Tartu Press.
- Lepajõe, Kersti. 2008. Personaalpronoomenite variandi valikust [About the choice of a personal pronoun variant]. *Lähivõrdlusi. Lähivertailuja* 18. 95–107. <http://dx.doi.org/10.5128/LV18.05>.
- Lepajõe, Kersti. 2011. *Kirjand kui tekstiliik. Riigieksamikirjandite tekstuaalsed, retoorilised ja diskursiivsed omadused [The Estonian high school matriculation examination essay as a text genre: textual, rhetorical and discursive features]*. (Dissertationes Philologiae Estonicae Universitatis Tartuensis 31). Tartu: University of Tartu Press.
- Levinson, Stephen C. 2006. Deixis. In Laurence R. Horn and Gregory Ward (eds.), *The handbook of pragmatics*, 97–121. Oxford: Blackwell Publishing Ltd.
- Levshina, Natalia. 2015. *How to do linguistics with R. Data exploration and statistical analysis*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Levshina, Natalia. 2017. A multivariate study of T/V forms in European languages based on a parallel corpus of film subtitles. *Research in Language* 15(2). 153–172. <https://doi.org/10.1515/rela-2017-0010>.
- Lindström, Liina. 2001. Verb-initial clauses in narrative. In Mati Ereht (ed.), *Estonian: Typological Studies V*, (Publications of the Department of Estonian of the University of Tartu 18.), 138–168. Tartu: Tartu University.
- Lindström, Liina. 2005. *Finiitverbi asend lauses. Sõnajärg ja seda mõjutavad tegurid suulises eesti keeles [The position of the finite verb in a clause: Word order and the factors affecting it in spoken Estonian]*. (Dissertationes Philologiae Estonicae Universitatis Tartuensis 16). Tartu: University of Tartu Press.
- Lindström, Liina. 2013. Between Finnic and Indo-European: Variation and change in the Estonian experiencer-object construction. In Leonid Kulikov (ed.), *The diachronic typology of non-canonical subjects*, 141–164. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Lindström, Liina. 2017. Lause infostruktuur ja sõnajärg. In Mati Ereht and Helle Metslang (eds.), *Eesti keele süntaks [Syntax of Estonian]*. (*Eesti keele varamu III.*), 537–656. Tartu: University of Tartu Press.
- Lindström, Liina, Mervi Kalmus, Anneliis Klaus, Liisi Bakhoff and Karl Pajusalu. 2008. Ainsuse 1. isikule viitamine eesti murretes. *Emakeele Seltsi Aastaraamat* 54. 159–185.
- Lindström, Liina, Maarja-Liisa Pilvik and Helen Plado. 2018. Nimetamiskonstruktsioonid eesti murretes: murdeerinevused või suuline süntaks? *Mäetagused* 70. 91–126. [https://doi.org/10.7592/mt2018.70.lindstrom\\_pilvik\\_plado](https://doi.org/10.7592/mt2018.70.lindstrom_pilvik_plado).
- Lindström, Liina and Virve-Anneli Vihman. 2017. Who needs it? Variation in experiencer marking in Estonian “need”-constructions. *Journal of Linguistics* 54. 1–34. <https://doi.org/10.1017/S0022226716000402>.
- Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.
- Lyons, John. 1977. *Semantics. Volume 2*. Cambridge: Cambridge University Press.
- Mäkinen, Leena, Soile Loukusa, Lea Nieminen, Eeva Leinonen and Sari Kunnari. 2014. The development of narrative productivity, syntactic complexity, referential cohesion and event content in four- to eight-year-old Finnish children. *First Language* 34(1). 24–42. <https://doi.org/10.1177/0142723713511000>.
- Metslang, Helena. 2012. On the case-marking of existential subjects in Estonian. *SKY Journal of Linguistics* 25. 151–204.

- Metslang, Helena. 2013. Coding and behaviour of Estonian subjects. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics* 4(2). 217–293. <https://doi.org/10.12697/jeful.2013.4.2.12>.
- Mihkla, Karl, Lehte Rannut, Elli Riikoja and Aino Admann. 1974. *Eesti keele lause-õpetuse põhijooned*. Tallinn: Valgus.
- Monzoni, Chiara M. and Ritva Laury. 2015. Making referents accessible in multy-party interaction. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics* 6(2). 43–62. <https://doi.org/10.12697/jeful.2015.6.2.02>.
- Nahkola, Tiina, Maria Reile, Piia Taremaa and Renate Pajusalu. 2020. Space, contrast and joint attention: Demonstrative adverbs in Russian, Estonian and Finnish. In Åshild Næss, Anna Margetts and Yvonne Treis (eds.), *Demonstratives in discourse*, 243–271. Berlin: Language Science Press. <https://doi.org/10.5281/ZENODO.4055830>.
- Nordlund, Taru, Ritva Laury, Renate Pajusalu and Külli Habicht. 2013. Kielikontaktit, kieliopillistuminen ja kieltenvälinen vertailu. In Leena Kolehmainen, Matti Miestamo and Taru Nordlund (eds.), *Kielet toistensa peilissä. Kielten vertailu tutkimusmenetelmänä*, 251–290. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Norris, Mark. 2018. Nominal structure in a language without articles: The case of Estonian. *Glossa: a journal of general linguistics* 3(1). 41. <https://doi.org/10.5334/gjgl.384>.
- Nuyts, Jan. 2007. Cognitive linguistics and functional linguistics. In Dirk Geeraerts and Hubert Cuyckens (eds.), *The Oxford handbook of cognitive linguistics*, 543–565. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199738632.013.0020>.
- Obana, Yasuko. 2003. Anaphoric choices in Japanese fictional novels: The discourse arrangement of noun phrases, zero and third person pronouns. *Text* 23(3). 405–443. <https://doi.org/10.1515/text.2003.017>.
- Oh, Sun-young. 2006. English zero anaphora as an interactional resource II. *Discourse Studies* 8(6). 817–846. <https://doi.org/10.1177/1461445606067332>.
- Pajusalu, Renate. 1996. Pronoun systems of common Estonian and Estonian dialects in a contrastive perspective. In Mati Ereht (ed.), *Estonian: Typological Studies I*, (Publications of the Department of Estonian of the University of Tartu 4.), 145–164. Tartu: University of Tartu.
- Pajusalu, Renate. 1997a. Is there an Article in (Spoken) Estonian? In Mati Ereht (ed.), *Estonian Typological Studies II*, (Publications of the Department of Estonian of the University of Tartu 8.), 146–177. Tartu: University of Tartu.
- Pajusalu, Renate. 1997b. Eesti pronoomenid I/1. Ühiskeele *see*, *too*, ja *tema/ta* [Estonian pronouns I/1. *see*, *too*, and *tema/ta* in common language. *Keel ja Kirjandus* 1. 24–30.
- Pajusalu, Renate. 1997c. Eesti pronoomenid I/2. Ühiskeele *see*, *too* ja *tema/ta* [Estonian pronouns. *see*, *too*, and *tema/ta* in common language]. *Keel ja Kirjandus* 2. 106–115.
- Pajusalu, Renate. 1998. Eesti pronoomenid II. Võru *sjoo*, *taa*, *tuu* ja *timä* [Estonian pronouns ii. Võru *sjoo*, *taa*, *tuu* and *timä*]. *Keel ja Kirjandus* (3). 159–172.
- Pajusalu, Renate. 1999. *Deiktikud eesti keeles [Deictics in Estonian]*. (Dissertationes Philologiae Estonicae Universitatis Tartuensiss 8). Tartu: University of Tartu Press.
- Pajusalu, Renate. 2000. Indefinite determiners *mingi* and *üks* in Estonian. In Mati Ereht (ed.), *Estonian: typological studies IV*, (Publications of the Department of Estonian of the University of Tartu 14.), 87–117. Tartu: University of Tartu Press.

- 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*. (*Studia Fennica Linguistica* 12.), 107–134. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Pajusalu, Renate. 2006a. Death of a demonstrative: person and time. The case of Estonian too. *Linguistica Uralica* XLII(4), 241–253.
- Pajusalu, Renate. 2006b. Multiple motivations for meaning of an elative wh-construction in Estonian. *Trames Journal of the Humanities and Social Sciences* 10(4), 324–340.
- Pajusalu, Renate. 2009. Pronouns and reference in Estonian. *Sprachtypologie und Universalienforschung* 62(1/2), 122–139. <https://doi.org/10.1524/stuf.2009.0008>.
- Pajusalu, Renate. 2015. Võro demonstratives: changing or disappearing? *Eesti ja soomeugri 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. 2017a. Nimisõnafraas [Noun phrase]. In Mati Ereht and Helle Metslang (eds.), *Eesti keele süntaks [Estonian syntax]*. (*Eesti keele varamu III.*), 379–404. Tartu: University of Tartu Press.
- Pajusalu, Renate. 2017b. Viiteseosed [Referential relations]. In Mati Ereht and Helle Metslang (eds.), *Eesti keele süntaks [Estonian syntax]*. (*Eesti keele varamu III.*), 566–589. Tartu: University of Tartu Press.
- Pajusalu, Renate, Helen Hint, Maria Reile and Piia Taremaa. 2020. Eesti keele viitevahendid kirjeldustes ja jutustustes [Influence of the context on the choice of referential devices in Estonian]. *Keel ja Kirjandus* 4, 313–335.
- Pajusalu, Renate, Maria Reile, Helen Hint, Tiina Nahkola and Piia Taremaa. 2018. Relative clauses in spatial and narrative contexts in Estonian, Finnish, and Russian. *SKY Journal of Linguistics* 31, 107–142.
- Pavlenko, Aneta. 2008. Narrative analysis. In Li Wei and Melissa G. Moyer (eds.), *The Blackwell guide to research methods in bilingualism and multilingualism*, 311–325. Blackwell Publishing Ltd. <https://doi.org/10.1002/9781444301120.ch18>.
- Payne, J. R. 2006. Noun phrase. In Keith Brown (ed.), *Encyclopedia of language and linguistics. Second Edition*, Vol. 5, 712–720. Elsevier Science.
- Pekelis, Olga E. 2018. Expletives, referential pronouns and pro-drop: The Russian extraposition pronoun *eto* in light of the English *it* and the German *es*. *Lingua* 203, 66–101. <https://doi.org/10.1016/j.lingua.2017.10.007>.
- Pook, Hanna. 2019. The pronoun *kes* “who” and its referent’s animacy in Estonian dialects. *SKY Journal of Linguistics* 32, 105–144.
- Pool, Raili. 1999. About the use of different forms of the first and second person singular personal pronouns in Estonian cases. In Mati Ereht (ed.), *Estonian: Typological Studies III*, (Publications of the Department of Estonian of the University of Tartu 11.), 158–184. Tartu: Tartu University.
- Prince, Ellen F. 1992. The ZPG letter: subjects, definiteness, and information-status. In Sandra A. Thompson and William C. Mann (eds.), *Discourse description: diverse analyses of a fund raising text*, 295–325. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Puolakainen, Tiina. 2015. Anaphora resolution experiment with CG rules. In *Proceedings of the Workshop on “Constraint Grammar – methods, tools and applications” at NODALIDA 2015, May 11–13, Vilnius, Lithuania*.
- R Core Team. 2020. R: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <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). 137–166. <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. 2019. *Estonian demonstratives in exophoric use: an experimental approach*. (Dissertationes Linguisticae Universitatis Tartuensis 34). Tartu: University of Tartu Press.
- Reile, Maria, Helen Plado, Harmen B. Gudde and 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 and 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/dx.doi.org/10.3176/lu.2019.3.02>.
- Reilly, Judy, Molly Losh, Ursula Bellugi and Beverly Wulfeck. 2004. “Frog , where are you?” narratives in children with specific language impairment, early focal brain injury, and Williams syndrome. *Brain and Language* 88. 229–247. [https://doi.org/10.1016/S0093-934X\(03\)00101-9](https://doi.org/10.1016/S0093-934X(03)00101-9).
- Rocca, Roberta, Kristian Tylén and Mikkel Wallentin. 2019. *This shoe, that tiger*: Semantic properties reflecting manual affordances of the referent modulate demonstrative use. (Ed.) Søren Wichmann. *PLOS ONE* 14(1). e0210333. <https://doi.org/10.1371/journal.pone.0210333>.
- Rocca, Roberta and Mikkel Wallentin. 2020. Demonstrative reference and semantic space: A large-scale demonstrative choice task study. *Frontiers in Psychology* 11. 629. <https://doi.org/10.3389/fpsyg.2020.00629>.
- Sacks, Harvey, Emanuel A. Schegloff and Gail Jefferson. 1974. A simplest systematics for the organization of turn taking for conversation. *Language* 50. 696–735. <https://doi.org/10.1016/b978-0-12-623550-0.50008-2>.
- Sahkai, Heete. 2003. Demonstrative doubling in spoken Estonian. *Trames* 7(2). 120–144.
- Sauermann, Antje and Natalia Gagarina. 2017. Grammatical role parallelism influences ambiguous pronoun resolution in German. *Frontiers in Psychology* 8. 1205. <https://doi.org/10.3389/fpsyg.2017.01205>.
- Schmitz, Katrin, Laura Di Venzio and Anna-Lena Scherger. 2016. Null and overt subjects in Italian and Spanish heritage speakers in Germany. *Lingua* 180. 101–123. <https://doi.org/10.1016/j.lingua.2016.04.004>.
- Scott, Kate. 2013. Pragmatically motivated null subjects in English: A relevance theory perspective. *Journal of Pragmatics* 53. 68–83. <https://doi.org/10.1016/j.pragma.2013.04.001>.
- Scriver, Olga and Manuel Díaz-Campos. 2016. Language variation suite: A theoretical and methodological contribution for linguistic data analysis. *Proceedings of the Linguistic Society of America* 1. 29. <https://doi.org/10.3765/plsa.v1i0.3734>.
- Searle, John R. 1969. *Speech acts: An essay in the philosophy of language*. Cambridge, UK: Cambridge University Press.
- Seppänen, Eeva-Leena. 1998. *Läsnäolon pronominit: tämä, tuo, se ja hän viittaamassa keskustelun osallistujaan*. Helsinki: Suomalaisen Kirjallisuuden Seura.

- Serratrice, Ludovica. 2013. The role of number of referents and animacy in children's use of pronouns. *Journal of Pragmatics* 56(1). 31–42.  
<https://doi.org/10.1016/j.pragma.2013.05.008>.
- Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press.
- Siiman, Ann. 2018. Variation of the Estonian singular long and short illative form: A multivariate analysis. *SKY Journal of Linguistics* 31. 139–167.
- Silvennoinen, Olli O. 2018. Constructional schemas in variation. Modelling contrastive negation. *Constructions and Frames* 10(1). 1–37.  
<https://doi.org/10.1075/cf.00009.sil>.
- Silverstein, Michael. 1976. Shifters, verbal categories, and cultural description. In K.H. Basso and H.A. Selby (eds.), *Meaning in anthropology*, 11–55. Albuquerque: University of New Mexico Press.
- Stefanowitsch, Anatol. 2008. Negative entrenchment: A usage-based approach to negative evidence. *Cognitive Linguistics* 19(3). 513–531.  
<https://doi.org/10.1515/COGL.2008.020>.
- Stein, Nancy and Christine Glenn. 1979. An analysis of story comprehension in elementary school children. In Roy O. Freedle (ed.), *New Directions in Discourse Processing*, Vol. 2, 53–120. Norwood, NJ: Ablex.
- Stevenson, Rosemary J., Rosalind A. Crawley and David Kleinman. 1994. Thematic roles, focus and the representation of events. *Language and Cognitive Processes* 9. 473–592.
- Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin and Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC bioinformatics* 9. 307. <https://doi.org/10.1186/1471-2105-9-307>.
- Strobl, Carolin, James Malley and 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>.
- Szmrecsanyi, Benedikt and Lars Hinrichs. 2008. Probabilistic determinants of genitive variation in spoken and written English: A multivariate comparison across time, space, and genres. In Terttu Nevalainen, Irma Taavitsainen, Päivi Pahta and Minna Korhonen (eds.), *The dynamics of linguistic variation: corpus evidence on english past and present*, 291–309. Amsterdam: John Benjamins Publishing Company.
- Taboada, Maite and Loreley Hadic Zabala. 2008. Deciding on units of analysis within Centering Theory. *Corpus Linguistics and Linguistic Theory* 4(1). 63–108.  
<https://doi.org/https://doi.org/10.1515/CLLT.2008.003>.
- Taboada, Maite and Loreley Wiesemann. 2010. Subjects and topics in conversation. *Journal of Pragmatics* 42(7). 1816–1828.  
<https://doi.org/10.1016/j.pragma.2009.04.009>.
- Tael, Kaja. 1988. Infostruktuur ja lauseliigendus [Information structure and the sentence structure]. *Keel ja Kirjandus* 3. 133–143.
- Tagliamonte, Sali A. and 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>.
- Takahashi, Daiko. 2008. Quantificational null objects and argument ellipsis. *Linguistic Inquiry* 39(2). 307–326. <https://doi.org/10.1162/ling.2008.39.2.307>.
- Talmy, Leonard. 2017. *The targeting system of language*. Cambridge/Massachusetts: The MIT Press.

- Tammekänd, Liina. 2015. Demonstratives in Võro and Estonian oral narratives. *Eesti ja soome-ugri keeleteaduse ajakiri* / The Journal of Estonian and Finno-Ugric Linguistics 6(2). 191–216. <https://doi.org/10.12697/jeful.2015.6.2.08>.
- Tao, Liang and Alice F. Healy. 2005. Zero anaphora: Transfer of reference tracking strategies from Chinese to English. *Journal of Psycholinguistic Research* 34(2). 99–131. <https://doi.org/10.1007/s10936-005-3634-5>.
- Taremaa, Piia. 2017. *Attention meets language: a corpus study on the expression of motion in Estonian*. (Dissertationes Linguisticae Universitatis Tartuensis 29). Tartu: University of Tartu Press.
- Taremaa, Piia, Helen Hint, Maria Reile and Renate Pajusalu. 2021. Constructional variation in Estonian: Demonstrative pronouns and adverbs as determiners in noun phrases. *Lingua* 254. 103030. <https://doi.org/10.1016/j.lingua.2021.103030>.
- Tauli, Valter. 1980. *Eesti grammatika. II: Lauseõpetus [Estonian grammar . II: Syntax]*. Uppsala: Finsk-ugriska institutionen.
- Tirkkonen, Mari-Epp. 2007. Ranniku- ja kirdemurde endofoorsed pronoomenid [Endophoric pronouns in Northeastern coastal dialect]. *Keel ja Kirjandus* 4. 277–291.
- Tognini-Bonelli, Elena. 2011. *Corpus linguistics at work*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Ueno, Mieko and Andrew Kehler. 2016. Grammatical and pragmatic factors in the interpretation of Japanese null and overt pronouns. *Linguistics* 54(6). 1165–1221. <https://doi.org/10.1515/ling-2016-0027>.
- Verhagen, Arie. 2019. Shifting tenses, viewpoints, and the nature of narrative communication. *Cognitive Linguistics* 30(2). 351–375. <https://doi.org/10.1515/cog-2018-0058>.
- Vihman, Virve-Anneli. 2015. Pick it up: A look at referential devices in Estonian child-directed speech. *Eesti ja soome-ugri keeleteaduse ajakiri* / *Journal of Estonian and Finno-Ugric Linguistics* 6(2). 63–83. <https://doi.org/10.12697/jeful.2015.6.2.03>.
- Vihman, Virve-Anneli and George Walkden. 2021. Verb-second in spoken and written Estonian. *Glossa: a journal of general linguistics* 6(1). 15. <https://doi.org/10.5334/gjgl.1404>.
- Vija, Maigi and Renate Pajusalu. 2009. Pronouns in Estonian child language. *Ad verba liberorum: Linguistics & Pedagogy & Psychology* 1(1). 23–31.
- Vilkuna, Maria. 1992. *Referenssi ja määräisyys suomenkielisten tekstien tulkinnessa*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Vliet, Sarah van. 2009. Reference points and dominions in narratives. A discourse level exploration of the referent point model of anaphora. In Vyvyan Evans and Stephanie Pourcel (eds.), *New directions in cognitive linguistics*, 441–464. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Vogels, Jorrig, Emiel Krahmer and Alfons Maes. 2012. Who is where referred to how, and why? The influence of visual saliency on referent accessibility in spoken language production. *Language and Cognitive Processes* 28(9). 1323–1349. <https://doi.org/10.1080/01690965.2012.682072>.
- Vogels, Jorrig, Emiel Krahmer and Alfons Maes. 2015. How cognitive load influences speakers' choice of referring expressions. *Cognitive Science* 39(6). 1396–1418. <https://doi.org/10.1111/cogs.12205>.
- Vogels, Jorrig, Emiel Krahmer and Alfons Maes. 2019. Accessibility and reference production: The interplay between linguistic and non-linguistic factors. In Jeanette K. Gundel and Barbara Abbott (eds.), *The Oxford handbook of reference*. Oxford, UK: Oxford University Press.

- Walker, Marilyn, Aravind K. Joshi and Ellen F Prince. 1998. Centering in naturally occurring discourse: an overview. In Marilyn Walker, Aravind K. Joshi and Ellen F Prince (eds.), *Centering theory in discourse*, 1–28. Oxford: Clarendon Press.
- Weir, Andrew. 2012. Left-edge deletion in English and subject omission in diaries. *English Language and Linguistics* 16. 105–129.  
<https://doi.org/10.1017/S136067431100030X>.
- Weiss, Daniel. 2004. The rise of an indefinite article: The case of Macedonian *eden*. In Walter Bisang, Nikolaus P. Himmelmann and Björn Wiemer (eds.), *What makes Grammaticalization?: a look from its Fringes and its Components*, 139–168. Berlin: Walter de Gruyter.
- Whaley, Lindsay J. 1997. *Introduction to typology: The unity and diversity of language*. Thousand Oaks, CA: SAGE Publications.
- Yurayong, Chingduang. 2020. *Postponed demonstratives in Finnic and North Russian dialects*. Helsinki: University of Helsinki.

## SUMMARY IN ESTONIAN

### Fraasist nullini: eesti keele viitamisvahendite mitmefaktoriline, vormispetsiifiline ja keeltevaheline analüüs

#### Teoreetiline taust

*Viitamine ehk referents.* Keelt kasutades ja omavahel suheldes peavad kõnelejad üldjuhul nimetama erinevaid tegevusi, seisundeid ja protsesse ning olendeid, asju, objekte ja nähtusi, kes või mis neis tegevustes osalevad või kellele need on suunatud. Kõneaktide teooria järgi jagunevad kõneaktid predikatiivseteks (väljendavad tegevusi) ja referentsiaalseteks (väljendavad entiteete) (Searle 1969; P. Hanks 2019). Siinse kognitiivse keeleteaduse ja funktsionalismi põhimõtetest lähtuva doktoritöö fookuses on referents ehk viitamine, st viis, kuidas kõnelejad osutavad kõnelemise käigus erinevatele entiteetidele. Entiteetidena käsitletakse mis tahes füüsilisi või mentaalseid üksusi, mis on ajas suhteliselt püsivad ning millele on võimalik keeleliselt viidata (nt inimesed ja loomad, objektid ja abstraktsed nähtused). Kui entiteedile on mingis kindlas lauses keelelise vahendiga viidatud, siis tekib diskursusesse referent (Lambrecht 1994: 37). Referendid on seega diskursuse elemendid, need luuakse ning need muutuvad konkreetsetes suhtlussituatsioonides (W. F. Hanks 1990; Cornish 1999; Etelämäki 2009).

Referentsiaalsusest rääkides eristatakse semantilist ja pragmaatilist referentsikäsitlust (Abbott 2010; vt ka Gundel ja Abbott 2019). Semantilist käsitlust iseloomustab seisukoht, et viitaval keelevahendil iseenesest on seos keelevälise entiteediga, millele see viitab. Sellisena on referentsi vaadelnud mitmed filosoofid, nt Frege, Wittgenstein, Russell, Carnap. Pragmaatilise käsitluse keskmes seevastu ei ole keeleväljend ise, vaid kõneleja. Selline pragmaatiline vaade on rohkem levinud keeleteadlaste seas, ka siinses töös, ning selle kohaselt on referents kolme osapooliga suhe, mis hõlmab kõnelejat, keelelist vahendit ning viidatavat referenti (Abbott 2010: 2). Ent kuna oma olemuselt on referents suhtlustegu, siis peab adekvaatne referentsiaalsust kirjeldav mudel kaasama ka teksti adressaadi(d). Seda seetõttu, et efektiivne viitevahendi valik peab võtma arvesse kuulaja teadmisi ja mentaalset staatust, vastasel korral ei saa viitesuhte loomine referendi ja keelevahendi vahel olla adressaadi jaoks tulemuslik (Ariel 1990, 2001; Chafe 1994; Gundel jt 2010, 1993; Hanks 1990; Kibrik 2011; Lambrecht 1994).

*Anafooriline vs. deiktiline viitamine.* Viitevahendi valikul on oluline roll ka käimasoleva kõnesituatsiooni kontekstil. Nimelt on võimalik viitamisel tugineda nii keelesisesele kui ka keelevälisele maailmale ning sellest lähtuvalt on viitavaid akte traditsiooniliselt jaotatud deiktilisteks (eksofoorilisteks, ruumilisteks) ja anafoorilisteks (endofoorilisteks, tekstilisteks) (Bühler 1990[1934]; Fillmore, 1997; Cornish, 1999; Halliday & Hasan, 1976; Kibrik, 2011; Lambrecht, 1994). Näiteks kui kõneleja hüüab „Vaata seda!“ ning osutab seda tehes näpuga ühe

suure ämbliku suunas, on tegu deiktilise viitamisega, sest viidatav entiteet on reaalselt kõnesituatsioonis kohal. Ent kui kõneleja ütleb „Eile oli mu vannitoas hiiglaslik ämblik. See tahtis mind ära süüa“, siis on see näide anafoorilisest viitamisest, kuna fraasiga *see* viidatud referent on kuulaja jaoks loodud käimasolevas kõnesituatsioonis ning kuulaja saab seda tuvastada vaid talle teada oleva keelelise konteksti põhjal.

Nüüdisaegsed käsitlused rõhutavad aga, et selge piiri tõmbamine anafoorilise ja deiktilise viitamise vahele ei ole võimalik ega ka otseselt vajalik (nt Laury 1997; Etelämäki 2009; Jarbou 2010; Talmy 2017). Näiteks leiab Talmy (2017), et nii anafoorilist kui deiktilist viitamist mõjutavad samad mentaalsed protsessid, ning et lisaks sellele ei ole referentsiaalse vahendi valikul alati määravaks mitte niivõrd kontekst, kuivõrd referendi kontseptualiseerimise viis, mis võib sõltuda konkreetsest vahendist endast, mitte niivõrd ümbritsevast situatsioonist. Siinses töös vaatluse all olevaid viitavaid fraase on nimetatud anafoorilisteks, et säilitada terminoloogiline selgus ja ühtsus teiste uurimustega. Seda tehes ei ole aga võetud seisukohta, et anafoorilist ja deiktilist viitamist peaks alati tingimata eraldi käsitlema. Pigem oleks edaspidi kasulik vaadelda just anafoorilise ja deiktilise viitamise omavahelist interaktsiooni, eriti kuna nii anafoorilise kui deiktilise viitamise jaoks kasutatakse paljudes keeltes sarnaseid keeleüksusi (nimisõna-fraase).

*Erinevad viitevahendid.* Referentidele osutamiseks kasutatakse keeles enamasti nimisõna-fraase, kusjuures nimisõna-fraasi struktuur võib olla väga varieeruv. Üldisel tasandil jagatakse referentsiaalsed nimisõna-fraasid tavaliselt täielikeks nimisõna-fraasideks (nt *üks hea raamat*, *õnn*) ning semantilise sisu poolest vaesteks pronoomeniteks (nt *see*, *ta*) (Abbott; Kibrik 2011; Pajusalu 2017a). Täpsemaid referentsiaalsete nimisõna-fraaside jaotusi nende struktuuri alusel on mitmeid (nt Ariel 1990; Givón 1983b; Gundel jt 1993), siinses töös on vaatluse all vaid kõige enam levinud põhilised nimisõna-fraasi tüübid. Toetudes eesti keele struktuuri akadeemilisele kirjeldusele (Erelt jt 1993; Erelt, Metslang 2017) ning varasematele uurimustele eesti keele referentsiaalsetest vahenditest, on doktoritöös lähema vaatluse alla võetud viis referentsiaalsete vahendite gruppi: i) artikli-laadse määratlejaga NPd, nii definiitsed (nt *see naine*) kui indefiniitsed (nt *üks maja*), ii) määratlejata täielikud NPd (nt *ratas*, *uus ratas*), iii) demonstratiivid (*see*, *too*, *siin*, *seal* jt), iv) kolmanda isiku personaalpronoomenid (*ta*, *tema*, *nad*, *nemad*) ning v) kolmanda isiku nullviitamine (*õ sõi putru*).

Nimisõna-fraasid võivad olla kas definiitsed või indefiniitsed, st nendega on võimalik väljendada referendi identifitseeritavust käimasolevas diskursuses (Chafe 1994; Chesterman 1991; Laury 2001; Lyons 1999). Definiitsuse kateooria võib avalduda grammatilisel tasandil, seda eelkõige keeltes, milles on olemas grammatilised artiklid (nt inglise *a child* ja *the kid*) (Dryer 2013b; 2013a). Pragmaatilisel tasandil on võimalik referendi tuntust või identifitseeritavust aga väljendada ka keeltes, milles grammatilist artiklit pole. Selleks kasutatakse tihti määratlejaid ehk funktsioonisõnu, mille eesmärk ongi väljendada NP informatsioonilist staatust ja seoseid teiste referentidega (Payne 2006; Pajusalu 2017a).

Eesti keeles on sagedamad määratlejad demonstratiivpronoomen *see* (*poiss elab selles majas*), indefiniitne pronoomen *üks* (*poiss rääkis ühe nalja*) ja pronoomen *oma* (*poiss võttis oma mütsi*), ent ka paljud teised pronoomenid (vt lähemalt Pajusalu 2017a; artiklid DETERMINERS ja DEMONSTRATIVES). Ent definiitsuse pragmaatilise olemuse tõttu on mõisteta, et loomulikus keeles esineb sageli ka selliseid viitavaid fraase, mis kontekstist lahutatuna ei võimalda üheselt otsustada, kas tegu on definiitse või indefiniitse viitamisega (Kibrik 2011; Laury 2001).

*Viitamisvahendi valikut mõjutavad tegurid.* See, millise viitamisvahendi valib kõneleja konkreetsetes kõnesituatsioonis, sõltub mitmest erinevast tegurist. Klassikaks saanud pragmaatiliste lähenemiste järgi on peamine ja tähtsaim viitava fraasi valikut mõjutav tegur seotud kõnelejate tähelepanu fookusega, mis avaldub diskursuses referendi esilduvusena (nt Givón 1983; Ariel 1990; Gundel jt 1993, Chafe 1994): mida esilduvam on referent, seda lühema vormiga on võimalik sellele referendile viidata. Esilduvusel põhinevad teooriad leiavad, et inimkognitsioonist lähtuvalt on üsna universaalne tendents, et n-ö vähem referentsiaalsed sisu evivad lühikesed vormid (nt nullviitamine ja personaalpronoomenid) osutavad kõige esilduvamatele diskursuse-referentidele, seevastu pikemad fraasid (nt täielikud nimisõnafrasid, eriti indefiniitsed, aga ka definiitsuse suhtes markeerimata NPd) osutavad referentidele, mis on diskursuses alles uued ja seega vähem esilduvad. Ka siinse doktoritöö esimeses artiklis (ZERO) lähtekohaks võetud tsenderdamisteooria (*Centering theory*, edaspidi CT) kuulub oma põhiolemuselt pragmaatilisel valikul põhinevate lähenemiste alla, ehkki teooria aluseks olevad reeglid ja kitsendused on ülimalt täpsed ja paindumatud (Grosz jt 1995; Walker jt 1998).

On aga hulgaliselt erinevaid tegureid, mis referendi esilduvust ja seega ka viitava fraasitüübi valikut mõjutavad. Sellisteks teguriteks on näiteks infostruktuur (nt Colonna jt 2012; Crawley 1986; Gernsbacher 1990; Järvikivi jt 2005), süntaktiline roll (Crawley, Stevenson ja Kleinman 1990; Grosz jt 1995), süntaktilise rolli parallelism (Chambers ja Smyth 1998; Sauermann ja Gagarina 2017), referendi temaatiline roll (Arnold 2001; Stevenson, Crawley ja Kleinman 1994), semantika ja üldine maailmateadmine (Hobbs 1979) ning mitmed koherentsussuhted (Kehler jt, 2008). Nende uurimuste kontekstis on mõisteta, miks vastukaaluks n-ö ühe faktori lähenemisele on mitmed uurijad viimasel ajal hakanud pooldama pigem n-ö mitmefaktorilist lähtekohta referentsiaalse valiku selgitamiseks (nt Kibrik 1996, 2011; Kaiser ja Trueswell 2008; Brown-Schmidt jt 2005). Mitmefaktorilise lähenemise tuumaks on idee, et viitevahendi valikut ei saa adekvaatselt seletada, kui lähtuda vaid referendi esilduvuse (või mistahes muu ühe teguri) mõjust, vaid et samal ajal tuleb arvesse võtta mitmeid erinevaid tegureid, nii grammatilisi, semantilisi, kui pragmaatilisi, ja nii referendi endaga, aga ka eelneva referentsiaalse fraasi või eelneva mainimise kaugusega seotud tunnuseid (Kaiser ja Trueswell 2008; Kibrik 2011; Kibrik jt 2016; Brown-Schmidt jt 2005). Lisaks sellele peaks üldine viitevahendi valikut selgitav mudel arvesse võtma, et erinevad vormid võivad olla seotud erinevate faktoritega, st

mõned olulised faktorid võivad seostuda vaid mõne viitevahendiga, samas kui teised faktorid võivad olla olulised üle kogu võimalike vahendite spektri (vt nt Kaiser ja Trueswell 2008).

Siinne doktoritöö võtab aluseks vormispetsiifilise mitmefaktorilise uurimisraamistiku ehk eelduse, et erinevaid viitevahendeid valides lähtub kõneleja erinevatest mõjuteguritest. Mõned tegurid võivad eri vahendite puhul kattuda, samas kui teised tegurid mõjutavad vaid mõne konkreetse viitevahendi valikut. Lisaks keelesisestele erinevustele vaatlen töös ka seda, kas ja kuidas referentsiaalset valikut mõjutavad tegurid eri keeltes kattuvad või erinevad. Doktoritöö lähtub ideest, et isegi kui eri keeltes on referentsiaalseid vahendeid, mis oma grammatilise struktuuri ja sisu poolest sarnanevad (nt personaalpronoomenid, demonstratiivpronoomenid), siis tegelikult keelekasutuses võib neil vormidel olla väga erinevaid funktsioone.

*Referentsiaalsed fraasid eesti keeles.* Eesti keele viitavaid fraase on süsteemselt uuritud alates 1990ndate teisest poolest, kusjuures uurimused on keskendunud erinevatele viitevahendite tüüpidele ning erinevatele keelekasutuskontekstidele, samuti on kasutatud erinevaid meetoodilisi lähenemisi. Kõige põhjalikumad käsitlused eesti keele viitevahendite kasutamise seaduspäradest pärinevad Renate Pajusalult (1999, 2005, 2006, 2009, 2017b; Pajusalu jt 2020). Mitmed uurijad on analüüsinud viitevahendite kasutuse eri aspekte ja nüansse, nt eesti keele määratlejate referentsiaalseid omadusi (Pajusalu 1997, 2000; Hiietam 2003), demonstratiivide funktsioone deiktilises (ruumilises) kasutuses (Reile 2015, 2016; Pajusalu jt 2018; Reile jt 2019; Reile jt 2020), personaal- ja demonstratiivpronoomeni vaheldust anafoorilises kontekstis (Kaiser ja Hiietam 2003; Kaiser ja Vihman 2010), kolmanda isiku lühikese ja pika pronoomeni vaheldust (Kaiser 2010), referentsiaalseid pronoomeneid murdekeeles (Pajusalu 2006, 2015; Tirkkonen 2007; Tammekänd 2015).

Võrreldes varasemate eesti viitevahendite uurimustega on siinsel doktoritööl mõnevõrra erinev fookus, sest töö ei keskendu ühe konkreetse viitevahendi funktsioonidele ja omadustele. Selle asemel on doktoritöö huviobjektiks, kuidas erinevad referentsiaalsed üksused diskursuses viiteahelaid moodustavad, st milliseid vahendeid kasutatakse diskursuses sarnastele referentidele viitamiseks. Seega ei ole fookus mitte ühele või teisele viitevahendile iseloomulikel referentsiaalsetel omadustel, vaid hoopis teguritel, millest sõltub see, millise referentsiaalse fraasi ja miks valib kõneleja mitme erineva võimaliku variandi seast.

## Töö eesmärgid

Doktoritööl on kolm olulist lähtekohta. Esiteks on tegu uurimusega, mis keskendub üksikasjalikult eesti keele viitevahendite süsteemile. Teiseks on doktoritöö huviorbiidis erinevused selliste viitamisvahendite vahel, millel on küll võrdlemisi sarnased funktsioonid, kuid mis ei ole siiski täissünonüümid. Kolmandaks on doktoritöös eesti keele viitevahendeid ja nende süsteemi kõrvutatud



vastavate süsteemidega teistes keeltes (soome ja vene), et näha, kuivõrd ja mille poolest erinevad eri keelte viitamisvahendite süsteemid üksteisest.

Doktoritöös olen seadnud neli eesmärki.

- 1) Välja selgitada, millised referentsiaalsed vahendid kuuluvad eesti keele referentsiaalsete vahendite põhisüsteemi, st milliste vahenditega moodustakse loomulikus diskursuses kõige tüüpilisemalt koherentseid viiteahelaid. (Artiklid ZERO, DETERMINERS, PRONOUNS, DEMONSTRATIVES)
- 2) Vaadata kaugemale pelgalt referendi esilduvuse (või tuntuse) olulisusest ning selgitada välja, millised referendi enda, sõna tähenduse või keelestruktuuriga seotud faktorid mõjutavad referentsiaalse vahendi valikut diskursuses. (Artiklid ZERO, DETERMINERS, PRONOUNS, DEMONSTRATIVES)
- 3) Analüüsida teatud referentsiaalsete vahendite paarikuid, mis eesti keeles alluvad konstruktsioonilise varieerumise põhimõtetele, ning selgitada neisse kuuluvate vahendite referentsiaalseid omadusi. Täpsemalt vaatlen doktoritöös, mille poolest erinevad i) kolmanda isiku lühike pronoomen *ta* ning sellega sisu poolest sarnane nullviitamine (artikkel ZERO) ning ii) määratlejana nimisõnafraasi koosseisus esinevad demonstratiivpronoomen ja demonstratiivverb (artikkel DEMONSTRATIVES).

Mõnevõrra sarnaste paaridena on vaatluse all ka erinevused iii) määratlejaga ning määratlejata nimisõnafraaside vahel (artikkel DETERMINERS) ja iv) pronoomenite ning täielike nimisõnafraaside vahel (artikkel PRONOUNS). Kuna kaks viimast fraasitüüpide paari ei ole käsitletavad konstruktsioonilise varieerumisena, on nende uurimuste teoreetilised lähtekohad mõnevõrra erinevad, kuid põhiküsimus referentsiaalsete funktsioonide olemusest neis uurimustes kattub.

- 4) Doktoritöös on oluline roll ka keeltevahelisel võrdlusel, et anda ülevaade sellest, kas ja mille poolest erinevad keelestruktuuris sarnasel kohal ja sarnaselt kirjeldatud referentsiaalsed fraasid oma referentsiaalsete omaduste ja funktsioonide poolest. Täpsemalt võrdlen selleks definiitsete ja indefiniitsete määratlejate kasutust eesti ja soome keeles (artikkel DETERMINERS) ning b) erinevate pronoomenite kasutusmustreid ja neid mõjutavaid faktoreid eesti, soome ja vene keeles (artikkel PRONOUNS).

Lisaks teoreetilisele küsimusepüstitusele käsitlen töös ka järgmisi metodoloogilisi probleeme.

- 1) Doktoritöös soovin välja töötada võimalikult lihtsat ja efektiivset andmekogumise metoodikat, mis võimaldab süsteemselt tutvustada keeltevahelisi erinevusi referentsiaalsete vahendite kasutuses. Selleks kasutan ja arendan töös pildiseeria põhjal suuliste narratiivide kogumise katset (artiklid DETERMINERS ja PRONOUNS).
- 2) Doktoritöö põhiliseks uurimismaterjaliks on terviklikud suulised narratiivid, sest referentsi uuritakse esmajoones diskursuse fenomenina ning mida

rohkem konteksti on uurijale kättesaadav, seda kindlamaid järeldusi saab ta teha (Kibrik 2011). Suulise keele kasuks otsustasin seetõttu, et narratiivide kogumise katses osalemine oleks katseisikutele võimalikult kiire ja lihtne. Siiski, siinse töö neljas uurimus (artikkel DEMONSTRATIVES) põhineb korpusel pärit materjalil, eesmärgiga näidata, et teatud uurimisküsimuste puhul on referentsiaalsete fraaside uurimine korpuslingvistika meetoditega asjakohane ning tulemuslik.

- 3) Töös rakendan erinevaid uurimisviise, mis võimaldavad analüüsida selliste sarnaseid vorme sisaldavate paaride liikmeid, milles i) üks on „nähtav“, st fonoloogiliselt realiseeruv, teine aga n-õ nullvorm, millel fonoloogilist väljendust ei ole (artikkel ZERO) või ii) üks esineb teatud keelekasutus-situatsioonis märgatavalt tihti, samas kui teine on selles situatsioonis väga harv, ent üldiselt keelekasutuses siiski esineb (artikkel DEMONSTRATIVES).
- 4) Siinne töö on teadaolevalt esimene uurimus, mis rakendab CT-s (vt Grosz jt 1995) välja töötatud analüüsiprotseduuri eesti keele andmestiku peal. Vaatlen, kas CT põhimõtted võimaldavad eristada ja selgitada erinevate referentsiaalsete vahendite kasutust ning kuivõrd on see teooria sobilik eesti keele materjalis esinevate seaduspärade uurimiseks (artikkel ZERO).

## Metoodika

Doktoritöös on kasutatud kaht tüüpi uurimismaterjali ning vaadeldud nii suulist kui kirjalikku keelekasutust. Kolmes artiklis (ZERO, DETERMINERS, PRONOUNS) on analüüsitud viitevahendite kasutust suulises keeles, nende uurimuste materjal on kogutud *katseliselt* vaba moodustuse katse käigus, kusjuures artiklites olen rakendanud kaht eri disaini. Katsetes osalejatel paluti ette antud stiimuli põhjal jutustada üks või mitu lugu. Töö esimeses artiklis (ZERO) on stiimulina kasutatud Pirnifilmi (Chafe 1980) ning katseisikutel paluti filmis nähtu ümber jutustada oma tuttavale, kellega koos nad katses osalesid. Kahe järgmise artikli (DETERMINERS, PRONOUNS) andmestik on aga kogutud kolme pildiseeria abil ning katseisikud jutustasid oma lugusid katse läbivijale, mitte oma tuttavale. Et võrrelda eesti keele viitevahendite süsteemi teiste keeltega, on pildiseeria alusel narratiive kogutud ka soome ja vene keele kõnelejatelt.

Doktoritöö viimane artikkel (DEMONSTRATIVES) tugineb *korpusmaterjalile*, mis pärineb Eesti keele ühendkorpusest 2017, st vaatluse all kirjalik keelekasutus. Kahe erinevat tüüpi materjali kombineerimine osutus vajalikuks, kuna uurimuse käigus selgus, et teatud referentsiaalsed fraasid esinevad teatud teksti-tüübis (suulistes narratiivides) liiga harva selleks, et teha nende referentsiaalsete omaduste kohta usaldusväärseid järeldusi. Tabelis 1 on kokkuvõtlikult esitatud doktoritöös kasutatud uurimismaterjali tüübid, uuritud keeled ja vastava materjali maht.

**Tabel 1.** Doktoritöös analüüsitud materjal

Uurimus	Uurimismaterjal	Keeled	Katseisikute arv / narratiivide arv	Analüüsitud NPde arv
<b>ZERO</b>	suuline narratiiv (filmi põhjal)	eesti	16/16	517
<b>DETERMINERS</b>	suuline narratiiv (pildiseeria põhjal)	eesti	20/59	676
		soome	20/54	680
<b>PRONOUNS</b>	suuline narratiiv (pildiseeria põhjal)	eesti	20/59	540
		soome	20/54	464
		vene	20/57	530
<b>DEMONSTRATIVES</b>	kirjalikud korpuslaused	eesti	-	2400

Iga doktoritöösse kuuluv uurimus vaatlleb mõnevõrra erinevaid seletavaid *tunnuseid*, mis võivad referentsiaalse vahendi valikut mõjutada. Täpsem tunnuste valik sõltub iga konkreetse uurimuse fookusest. Ülevaatlikult on kõik doktoritöös kasutatud selgitavad tunnused esitatud tabelis 2. Grammatilised tunnused on seotud fraasi morfoloogilise vormi või süntaktilise käitumisega, semantilised tunnused põhinevad NP peasõnaks oleva substantiivi tähendusel ning pragmaatilised tunnused näitavad NP käitumist diskursuses ning selle suhet teiste selle diskursuse NP-dega. Uuritavaks tunnuseks on igas uurimuses referentsiaalne valik ehk see, milline viitevahend teatud võimalike vahendite hulgast osutub konkreetsetes kasutussituatsioonis valituks.

**Tabel 2.** Uurimustes kasutatud seletavad tunnused

Seletava tunnuse tüüp	Tunnus	Artiklid
<b>CT-st lähtuv tunnus</b>	Ülemineku tüüp	ZERO
<b>Grammatilised tunnused</b>	Süntaktiline roll	ZERO,
	Kääne	DETERMINERS,
	Lausetüüp	PRONOUNS
	Arv	DETERMINERS, PRONOUNS
<b>Pragmaatilised tunnused</b>	Mainimiskord	DETERMINERS,
	Referentsiaalne kaugus	PRONOUNS
	Eelneva NP tüüp	PRONOUNS
<b>Semantilised tunnused</b>	Elusus	DETERMINERS,
	Teised elus referendid lauses	PRONOUNS
	Nimisõna semantiline klass (sh elusus)	DEMONSTRATIVES
	Nimisõna konkreetsus	
	Liikumisverb	

Analüüsitav uurimismaterjal on oma olemuselt mitmemõõtmeline ning hõlmab vaid kategoriaalseid tunnuseid, mistõttu on kvantitatiivseks andmeanalüüsiks kasutatud mitteparameetrilisi meetodeid: tingimuslikku rekursiivset otsustuspuid ehk klassifitseerimispuid (Hothorn jt 2006) ja tingimuslikku juhumetsa (Breiman 2001; Strobl jt 2008; Strobl jt 2009). Analüüside läbiviimiseks on kasutatud vabavaralist statistikatarkvara R (R Core Team 2021) pakette *party* (Strobl jt 2009; Hothorn jt 2015) tingimusliku otsustuspuid ja juhumetsa analüüsiks ning *sjPlot* (Lüdecke 2021) tabelite koostamiseks. Igas artiklis on esitatud ka kirjeldav statistika ning lisaks kvantitatiivsele analüüsile on välja toodud ka konkreetsete keelenäidete kvalitatiivseid selgitusi.

## Uurimuste tulemused ja arutelu

Kõik doktoritöös esitatud viitavate nimisõnafraaside analüüsid on üles ehitatud nii, et igas uurimuses kõrvutatakse selliseid viitevahendeid, mis on oma struktuurilt ja/või tähenduselt sarnased ning mida võib vähemalt teatud kontekstides üksteisega asendada. Esmalt on siinses peatükis nelja uurimuse kokkuvõttena kirjeldatud eesti keele viitevahendite süsteemi, sh eri viitevahendite referentsiaalseid omadusi ja nende valikut mõjutavaid tegureid. Sellele järgneb eesti keele viitevahendite referentsiaalset valikut määravate tegurite paralleelne võrdlus teiste keeltega, täpsemalt soome keelega (DETERMINERS) ning soome ja vene keelega (PRONOUNS).

### *Eesti keele viitevahendite süsteem ja valikupõhimõtted*

Artikkel ZERO keskendus eesti keele kolmanda isiku lühikese pronoomeni *ta* ning kolmanda isiku nullviitamise referentsiaalsetele omadustele ning uurimuse eesmärk oli välja selgitada kahe minimaalse viitefraasi erinevused, lähtudes CT (Grosz jt 1995; Walker jt 1998) analüüsiprotseduurist. Uurimismaterjaliks olid Pirnifilmi (Chafe 1980) põhjal jutustatud suulised narratiivid.

Pirnifilmi narratiivides esines nullviitamist veidi enam kui personaalpronoomeni kasutust (vastavalt 54% ja 46%). CT-s esitatud üleminekutüüpidele tuginedes ei ilmnenud siiski märkimisväärset erinevust selles, kuidas kaks viitevahendit diskursuse koherentsust mõjutavad, sest nii nullviitamine kui pronoomen olid ülekaalukalt seotud kõige sujuvama üleminekutüübiga (JÄTKAMINE). Täiendavate grammatiliste tunnuste (lausetüüp, kääne, süntaktiline roll) kaasamine analüüsi aga näitas, et keelestruktuuri tasandil on nullviitamine ja *ta* siiski erinevad: nullviitamine on iseloomulik lauses nominatiivse subjekti rollis olevate referentide korral ja see esineb enamasti pealauses. Personaalpronoomeni esinemisvõimalused ei ole kindlate grammatiliste kontekstidega piiratud ning pronoomenit *ta* kasutatakse vabalt nii nominatiivis kui teistes käänetes, nii subjektina kui mitte-subjektina ning lisaks pealausetele esineb see sageli ka kõrvallauses.

Artikli ZERO tulemused toetavad mitmefaktorilist ja vormispetsiifilist vaadet referentsiaalsele valikule, kuna pelgalt referendi esilduvusele või diskursuse

koherentsusele toetudes ei ole võimalik kahe vormi erinevusi selgitada. Et nullviitamine ja *ta* on enim seotud kõige esilduvamate referentidega, siis ei saa väita, et üks neist oleks esilduvuse skaalal teisest eespool. Niisiis on vajalik täiendavate tunnuste kaasamine, kuna kahe erineva keelelise vormi tähendus ja funktsioon, sh referentsiaalne tähendus ja funktsioon, ei saa olla täielikult sünonüümsed (Goldberg 1995, 2006; Janda ja Solovyev 2009). Siinne analüüs kinnitas, et grammatilised tunnused võimaldavad mõnevõrra nullviitamise kasutus-kontekste täpsustada.

Iseseisvate pronoomenite referentsiaalsed omadused olid ka artikli PRONOUNS fookuses. Lisaks nullviitamisele ja kolmanda isiku personaalpronoomenile käsitles see artikkel ka demonstratiivpronoomeneid ning demonstratiivadverbe. Artiklis PRONOUNS olid uurimismaterjaliks pildiseeria alusel kogutud lühinarratiivid, mis oma struktuurilt erinevad mõnevõrra Pirnifilmi narratiividest. Seetõttu on ka neis narratiivides kasutatud pronoomenite sagedusjaotus erinev. Uurimuse PRONOUNS andmestikus olid enim kasutatud pronoomeniteks personaalpronoomenid (52,8%), sageli kasutati ka nullviitamist (37,1%). Iseseisvaid anafoorilisi demonstratiivpronoomeneid ja -adverbe esines aga märgatavalt vähem (vastavalt 8,5% ja 1,6%).

Pronoomenite valikut eestikeelsetes suulistes narratiivides mõjutavad nii grammatilised kui semantilised ja pragmaatilised tegurid. Semantilised ja pragmaatilised tegurid, eelkõige elus referentide arv lauses ning mainimiskord, aga ka referentsiaalne distants, on seotud eelkõige demonstratiivpronoomeni valikuga. Demonstratiivpronoomenid esinevad enamasti elututele referentidele viitamisel (st kui lauses ei ole elusaid referente) nii referendi teisel kui ka hilisemal mainimisel. Personaalpronoomeni kasutamise määrab esmalt elusus kui semantiline tegur – personaalpronoomeniga viidatakse üldjuhul elusatele referentidele. Lisaks sellele eristub personaalpronoomen teistest pronoomenitest grammatiliste tegurite alusel (vt ka ZERO). Nullviitamist leidub personaalpronoomenist enam selliste pealause subjektide korral, millele viidati ka vahetult eelnevas lausungis (st referentsiaalne kaugus on 1). Statistiliselt olulisimaiks pronoomeni valikut määravateks teguriteks osutusid grammatilised tunnused: süntaktiline roll, lausetüüp ja kääne.

Eesti keele anafooriliste pronoomenite hulgas moodustavad ootuspäraselt omaette eristuva rühma demonstratiivadverbid, mis esiteks on viitevahenditena kõige harvemad ning teiseks seostuvad selgelt ruumitähenduste väljendamisega. Ka oma struktuuri ja sõnaliigilise kuuluvuse alusel on demonstratiivadverbid pronoomenitest lahus seisev klass, kuid kuna demonstratiivadverbid on sageli selgelt referentsiaalne funktsioon, siis on neid siinses käsitluses arvestatud teiste viitevahenditega samasse kategooriasse kuuluvana. Doktoritöö andmestikus leidis iseseisvaid referentsiaalsed demonstratiivadverbe siiski liiga vähe selleks, et nende referentsiaalsete omaduste kohta kindlaid järeldusi teha, mistõttu vajaks viitavad demonstratiivadverbid edasistes uurimustes põhjalikumat uurimist.

Seega võib järeldada, et eesti keele suulistes narratiivides eristuvad referentsiaalsete omaduste poolest enim personaalpronoomenid (lühikesed vormid) ja demonstratiivpronoomenid, samas kui personaalpronoomen ja nullviitamine on

referentsiaalse funktsiooni poolest sarnasemad. Nullviitamise võimalikud esinemiskontekstid on siiski piiratud. Kõige neutraalsem ja laialdasemate kasutusvõimalustega anafooriline pronoomen on eesti keeles personaalpronoomeni lühike vorm.

Doktoritöö artikkel DETERMINERS käsitles kõige sagedamini esinevaid artiklilaadseid määratlejaid, nii definiitseid kui indefiniitseid (nt *see* + NP, *üks* + NP), võrdlevalt eesti ja soome keele suulistes narratiivides. Artikli eesmärk oli välja selgitada, millised määratlejad mõlemas keeles suulistes narratiivides enim esinevad, kui sarnased või erinevad on eesti ja soome keeles kasutatavad määratlejad ning kas sarnaste määratlejate kasutust mõjutavad tegurid on eesti ja soome keeles samad või mitte.

Eesti keele andmestikus oli määretlejaga laiendatud 27,8% sellistest nimisõnafraasidest, millel on potentsiaal koos määratlejaga esineda. Määratlejatena olid eesti keeles enim kasutusel pronoomenid *see*, *üks*, *oma* ja *mingi*. Eesti keele määratlejate kasutust mõjutavad olulisemad tegurid on mainimiskord, süntaktiline roll ja elusus, samuti referentsiaalne kaugus. Olulisim on erinevus indefiniitsete (1. mainimiskord) ja definiitsete (2. mainimiskord) määratlejate vahel, neid gruppe omakorda mõjutavad eri tegurid. Indefiniitsete fraaside puhul määrab määratleja kasutust osaliselt elusus: elusate referentide nimetamisel esineb määratlejaid rohkem. Definiitsete määratlejate valikut mõjutab aga süntaktiline roll: subjektid ja objektid, kui neid on mainitud määratlejaga nimisõnafrasiga, saavad määratlejaks pronoomeni *see*.

Eesti keeles kasutatavate määratlejate esinemiskontekstid on kooskõlas universaalseks peetava tendentsiga, et teatud keelevahendid alluvad keeltes sageli grammatiseerimisele ning kaotavad järkjärgult oma semantilist sisu, liikudes üha grammatilisema tähenduse suunas (Heine jt 1991; Heine ja Kuteva 2006). Sellisteks keeleüksuseks on paljudes keeltes näiteks numeraal „üks“, mis grammatiseerub indefiniitseks artikliks (nt Givón 1981; C. Lyons 1999; Weiss 2004) ning demonstratiivpronoomenid, eriti anafoorilises kasutuses, mis on sageli grammatilise artikli lähtevormid (nt Greenberg 1978; Diessel 1999; Heine ja Kuteva 2006). Ent kuna eesti keeles ei ole määratlejad grammatilised, vaid selgelt alles pragmaatilised vahendid, siis ei saa (veel?) päriselt käsitleda eesti keele määratlejaid paralleelselt artiklikeelte artiklitega.

Artiklis DEMONSTRATIVES on vaatluse alla võetud sellised suulises narratiivis harva esinenud määratlejakonstruktsioonid, kus koos kohakäändes nimisõnafrasiga esineb määratlejana demonstratiivadverb (nt *siin*, *seal*). Neid konstruktsioone on võrdlevalt analüüsitud koos selliste määratlejakonstruktsioonidega, kus määratlejana esineb demonstratiivpronoomen (*see*, *too*). Uurimus on ajendatud tähelepanekust, et ka harvem esinevatel määratlejatel võib olla referentsiaalsuse seisukohast eristav funktsioon, ent uurimustes, mis keskenduvad üldisele süsteemile ja n-ö prototüüpsetele näidetele, jääb see sageli täpsemalt kirjeldamata.

Niisiis otsitakse töö viimases uurimuses vastust küsimusele, mille poolest erinevad teineteisest demonstratiivadverb ja demonstratiivpronoomen, kui need esinevad määratlejana koos kohakäändes nimisõnafrasiga (vrd *siin koolis* ja

*selles koolis*). Kuna suulistes narratiivides esines demonstratiivadverbist määratlejaga konstruktsioone liiga vähe selleks, et teha järeldusi selle määratleja referentsiaalsete funktsioonide kohta, siis on artikli DEMONSTRATIVES jaoks kogutud uurimismaterjal eesti keele ühendkorpusest 2017. Analüüsi kolme semantilise faktori (substantiivi konkreetlus, substantiivi semantiline klass ning kas klausis kasutatav verb on liikumisverb või mitte) rolli määratlejakonstruktsiooni valikul. Selgus, et demonstratiivse määratleja valik on tõesti seotud eelkõige substantiivi semantiliste faktoritega. Demonstratiivadverb esineb sagedamini koos kohta mitte tähistavate substantiividega, demonstratiivpronoomen aga just kohta väljendavate substantiividega. Kui tegu ei ole kohaga, siis mõjutab määratleja valikut omakorda substantiivi konkreetlus. Seega sai kinnitust oletus, et määratleja valik ise võib mõjutada seda, mil moel luuakse diskursusesse mingi referent.

Doktoritöö tulemused on niisiis kooskõlas töö lähte-eeldusega, et vormispetsiifiline mitmefaktoriline lähenemine võimaldab selgitada viitevahendite kasutusmehhanisme keeles adekvaatsemalt kui vaid referendi esilduvusel põhinevad teooriad. Töö raames läbi viidud uurimused illustreerivad hästi, et erinevate viitavate fraaside valik ja kasutus on mõjutatud erinevatest teguritest, kusjuures need tegurid võivad kuuluda nii grammatika, semantika kui pragmaatika tasandile. Ülevaade sellest, millised tunnused täpsemalt milliste viitefraaside kasutamist eesti keeles suunavad, on esitatud tabelis 3. Selgub, et vähemalt eesti keeles ei ole vaadeldud tunnuste seas ühte sellist, mis seostuks mingil moel kõikide paradigmasse kuuluvate viitevahenditega. Pigem on olukord vastupidine, st iga vahend on seotud omaette tunnuste komplektiga, mis ei kattu ühegi teise vahendi omaga täielikult. Näiteks süntaktiline roll omab kaalu nii nullviitamise, demonstratiivpronoomeni kui ka definiitsete määratlejaga fraaside valikul, ent teised tunnused, mis neid fraase samuti mõjutavad, on erinevad. Kui nullviitamise korral on lisaks süntaktilisele rollile määravad ka kääne ja referentsiaalne kaugus, siis nt demonstratiivpronoomeni valikut mõjutavateks tunnusteks on lisaks süntaktilisele rollile veel elusate referentide hulk lauses ning referendi mainimiskord.

**Tabel 3.** Eesti keele referentsiaalsete vahendite mitmefaktoriline vormispetsiifiline skeem

Mõjutavad faktorid	- süntaktiline roll, - kääne, - ref. kaugus, - (diskursuse koherents)	- ref. kaugus	- teised elusad referendid lauses, - süntaktiline roll, - mainimiskord	- kääne	- süntaktiline roll, - semantiline klass, - konkreetsus, - (mainimiskord ~ definiitne)	- elusus, - (mainimiskord ~ indefiniitne)
Esilduvus	Rohkem esilduv	>	>	>	>	Vähem esilduv
Vittevahend	nuliviitamine	pers-pron	dem-pron	dem-adv	def. detNP	NP indef. detNP
Kasutus-kontekstid	Nom subjekt pealauses	Subjekt ja mitte-subjekt nii pea- kui kõrvallauses	(enamasti) elutu subjekt	kohakäänded	( <i>see</i> NP, <i>demAdv</i> NP, <i>oma</i> NP)	( <i>üks</i> NP, <i>mingi</i> NP)  <i>iiks</i> NP – elusad ja elutud referendid  <i>mingi</i> NP – elutud referendid
		‘Neutraalne lühike viitevahend’				‘Neutraalne täielik viitevahend’



### *Keeltevaheline võrdlus*

Doktoritöö analüüs ei keskendunud vaid eesti keele viitevahendite süsteemile, vaid kõrvutas seda ka kahe lähema kontaktkeele – soome ja vene – viitamismustritega (artiklid DETERMINERS ja PRONOUNS). Lähteotuseks oli, et viitevahendite tegelikud funktsioonid võivad keeltes olla üsna erinevad, isegi kui viitevahendite inventar on võrdlemisi sarnane ning fraaside nimetused (nt demonstratiiv, personaalpronoomen jmt) keelekirjeldustes kattuvad (vt sarnast arutlust nt Haspelmath 2010). Uurimuse tulemused kinnitasid seda oletust.

Kolmes keeles on viitevahendite tüüpide sagedusjaotus suhteliselt sarnane – kõige enam esines määratlejata NPsid, neile järgnesid pronoomenid ning kõige vähem esines määratlejaga NPsid<sup>31</sup>. Suuremad erinevused ilmnesisid spetsiifilisemal viitevahendi tasandil. Ent fraasitüüpide esinemissagedused iseenesest ei selgita nende referentsiaalseid omadusi, seega on doktoritöös vaadeldud kõikide keelte puhul ka samu fraasi valikut mõjutavaid faktoreid. Võrreldes tingimusliku juhumetsa ja otsustuspuu analüüsi tulemusi kolmes keeles ilmneb, et viitevahendite valikut mõjutavad tunnused on keelte lõikes erinevad, st tunnused moodustavad keeltes eri viitevahenditega erinevaid kombinatsioone. Üldistatult võib öelda, et eesti keeles määravad määratlejaga fraasi valikut pigem pragmaatilised faktorid, soome keeles on suurem kaal aga grammatilistel faktoritel. Pronoomeni valikul seevastu on eesti keeles olulised just grammatilised faktorid, samas kui soome ja vene keeles on grammatilistel ja pragmaatilistel faktoritel pronoomeni valikule võrdset oluline mõju.

Demonstratiivpronoomenite kasutussagedus ning esinemiskontekstid on kolmes keeles üsnagi erinevad. Demonstratiivpronoomenite kasutuserinevused on osaliselt tingitud sellest, et keeliti varieerub demonstratiivide tüvede arv: kolm eri tüve soome keeles (*se, tämä, tuo*), kaks tüve vene keeles (*etot, tot*) ning üks või kaks eesti keeles (*see, mõnes piirkonnas see ja too*). Ent erinevused demonstratiivide kasutussageduses sõltuvad nähtavasti ka keelespetsiifilistest referentsiaalsete omaduste erinevustest, isegi siis, kui demonstratiivide tüvede arv on sarnane. Sellist erinevust võib kujutada kontinumina, mille ühes otsas asub vene keel, milles iseseisvaid demonstratiive kasutatakse võrreldes teiste pronoomenitega harva (1,5%). Kontinumi teises otsas paikneb aga soome keel, milles iseseisvad demonstratiivpronoomenid esinevad sageli (25%). Eesti keel on demonstratiivpronoomenite kasutussageduse poolest vene ja soome keele vahel, jäädes oma 8,5% kasutussagedusega siiski lähemale vene keelele. Samas ühendab eesti ja soome keelt eelistus viidata demonstratiivpronoomenitega elutele entiteetidele, ent vene keele materjalis leidunud väheste demonstratiivpronoomenitega osutati elusatele referentidele.

Kolmes keeles on aga sarnane personaalpronoomeni ja nullviitamise kasutavahekord. Kõige neutraalsem anafooriline viitevahend doktoritöö andmestikus on nii eesti, soome kui vene keeles personaalpronoomen, kuna see esineb väga sageli ning sellel ei ole väga ranged kontekstipiiranguid. Näiteks saab personaalpronoomen esineda eri käändevormides ning eri lausemallides, samuti võib

<sup>31</sup> Ei eesti, soome ega vene keeles esine grammatilist artiklit.

referentsiaalne kaugus eelmise mainimisega olla rohkem kui üks lausung. Nullviitamise eelduseks on see, et referenti on eelnevas (osa)lauses juba mainitud, st referentsiaalne kaugus eelmise mainimisega peab olema väike. Lisaks sellele ei esine nullviitamine tõenäoliselt eri käändeid kasutavates konstruktsioonides ning see on väga harv kõrvallausetes.

Ka määratlejana esineda võivate pronoomenite hulk on keeltes erinev. Määratlejaga fraaside osakaal on võrreldes teiste keeltega suurim eesti keeles, väga harva esines neid aga vene keele materjalis. Määratlejatena esinevad kõigis kolmes keeles kõige sagedamini demonstratiivpronoomenid definiitse määratleja rollis (nt ee *see poiss*, sm *se poika*, vn *etot mal'chik*), samas kui indefiniitsed määratlejad on omased vaid eesti materjalile.

Artiklis DETERMINERS läbi viidud analüüs näitas, et soome keele andmestikus oli määratlejate osakaal andmestikus eesti keelega võrreldes väiksem (17,8%). Sagedamini esinevateks määratlejateks osutusid soome keeles *se*, *tämä* ja possessiivsufiks. Soome keeles mõjutavad määratleja valikut kääne, mainimiskord ning elusus. Erinevalt eesti keelest eristuvad soome keeles n-ö muud, st harvem esinevad määratlejad, mis enamasti esinevad mitte-nominatiivis. Nagu eesti keele kõnelejad, eristavad ka soome keele kõnelejad definiitseid ja indefiniitseid määratlejaid. Soome keeles aga mõjutab elususe kategooria määratleja valikut definiitsete fraaside korral. Erinevalt eesti keelest ei näi soome keeles pronominaalsete määratlejate grammatiseatsioon artikliteks, vähemalt mitte indefiniitse artikli grammatiseatsioon, olevat veel eesti keelega võrreldes samal tasemel.

Niisiis sai uuringus üldiselt kinnitust, et i) sarnase grammatilise struktuuri ja nimetusega keeleüksustel on eri keeltes erinevad referentsiaalsed funktsioonid ning ii) neid referentsiaalseid funktsioone omakorda mõjutavad erinevad tunnused või iii) on sarnastel tunnustel mõju erinevatele viitevahenditele. Nende üldistuste taustal on oluline välja tuua kolm aspekti. Esiteks, kuna eri keeltes tulevad esile kindlad n-ö eelistatud anafoorilised viitevahendid, siis kinnitab see, et eri keelte kõnelejad toetuvad erinevatele referentsiaalsetele praktikatele (W. F. Hanks 1990), mida iseloomustavad ja kujundavad keele grammatiline süsteem, eri vahendite kasutussagedus üldiselt ning kõnesituatsioon ja kontekst. Teiseks selgub analüüsist, et vaadeldud keeled on artiklilaadsete määratlejate kasutuse näitel grammatiseatsiooniprotsessi eri staadiumitel, või et konkreetse fenomeni seisukohalt ei ole grammatiseatsioon mingis keeles (veel) oluline protsess (vt Heine ja Kuteva 2006). Grammatiseatsiooniprotsessis on aga muuhulgas oluline roll ka keelekontaktidel, mistõttu võib eeldada, et lähedastes keeltes võib mingil ajahetkel täheldada sarnaseid grammatiseatsiooniilminguid. Kolmandaks on oluline rõhutada, et referentsiaalne valik ei ole kunagi kategoriaalne ning alati tuleb viitevahendi valiku põhjuste kirjeldamisel arvestada kõneleja isiklike eelistustega, mida ei ole võimalik kindlate reeglite, põhimõtete või faktorite toel selgitada ega süsteemi paigutada (vt ka Kibrik 2011; Kibrik jt 2016). Kõnelejal on alati võimalik valida mitme eri viitevahendi vahel.

### *Piirangud*

Doktoritöö tulemuste tõlgendamisel tuleb arvesse võtta ka teatud piiranguid. Esiteks, analüüsitud referentsiaalset valikut mõjutavate tunnuste komplekt ei ole ilmselt ammendav ning on tõenäoline, et leidub veel mitmeid faktoreid, mis erinevaid viitevahendeid eri moel mõjutada võivad. Näiteks on leitud, et ka register või žanr, suhtlustegevus ning kõnelemise kontekst võivad määrata fraasi valikut (nt Pajusalu jt 2020; Taremaa jt 2021). Arvutipõhise referentsiaalse valiku ennustamise uurimisel on kodeeritud lausa üle 20 eri tunnuse (Kibrik jt 2016). Seega ei tohiks doktoritöö tulemusi pidada n-ö lõplikeks. Teiseks ei olnud võimalik kogutud uurimismaterjali põhjal vaadelda kõikide võimalike viitevahendite omadusi eestikeelses diskursuses. Seda seetõttu, et teatud fraasitüübid, nt pikad kolmanda isiku pronoomenid (*tema*) esinesid vaid paaril üksikul korral. Seega tuleks järgnevates uurimustes lähema vaatluse alla võtta just sellised fraasitüübid, mis praeguseks uuritud materjalis ei esine. Osalt on siinses töös sellega algust tehtud artiklis DEMONSTRATIVES, milles vaadeldi justnimelt suulistes narratiivides harva esinenud adverbilise määratlejaga NPsid. Samas ei ole täit selgust mitme viitevahendi referentsiaalsete omaduse osas, nt tuleks vaadelda harvem esinevate määratlejate referentsiaalseid omadusi (nt *mingi*NP), samuti ei ole praeguseks teada, kas ja mille poolest on referentsiaalselt erinevad erineva struktuuriga täielikud nimisõnafraasid (nt *väike poiss*, *see väike poiss*, *väike jooksev poiss* jne).

### *Metodoloogiline arutelu*

Lisaks teoreetilisele perspektiivile panustab siinne doktoritöö ka referentsiaalsuse uurimise metodoloogia arendamisse. Lähemalt on töös arutletud kahe metodoloogilise aspekti üle: i) referentsiaalsete fraaside uurimiseks sobiva andmestiku kogumine ning ii) referentsiaalsete fraaside analüüsi tarbeks kodeerimisega seotud otsused.

i) Andmestiku kogumiseks on töös kasutatud kaht eri lähenemist. Esimesed kolm artiklit (ZERO, DETERMINERS, PRONOUNS) tuginevad katseliselt kogutud suulistele narratiividele, mis on traditsiooniliselt olnud viitevahendite uurimisel üks eelistatuim andmestiku tüüp (nt Chafe 1980; Du Bois 1987; Berman ja Slobin 1994; Hickmann 2004; Hedberg 2010; Serratrice 2013; Disbray 2016; Verhagen 2019; Dancygier 2019). Katseliselt narratiive kogudes on oluline silmas pidada seda, millist mõju võib katse disain avaldada keelematerjali kvaliteedile. Näiteks siinses töös on kasutusel kaks erinevat disaini – lühikese filmiklipi (Pirnifilm, vt Chafe 1980) alusel narratiivide kogumine ning pildiseeriade alusel narratiivide kogumine. Mõlemal disainil on oma eelised teise ees. Näiteks on pildiseeriade põhjal narratiive lihtsam koguda, samuti on pildiseeriates ilmselt lihtsam piirata nende referentide arvu, keda katseisik oma narratiivis mainib. Mida ühetaolisematele referentidele katseisikud oma narratiivides osutavad, seda adekvaatsemalt saab võrrelda erinevaid viitevahendeid. Filmiklipi alusel kogutud narratiivid osutusid siinses töös aga mõnevõrra loomulikemateks, kuna katsesituatsioon sarnanes n-ö reaalse elu situatsioonile rohkem.

Doktoritöö neljas artikkel (DEMONSTRATIVES) seevastu vaatleb korpusest kogutud kirjalikku keelematerjali. Korpusandmestik ei ole viitesuhteid käsitlevates uurimustes kuigi levinud, kuid nagu siinses töös selgus, siis osutub korpusmaterjal väärtuslikuks selliste fraaside uurimisel, mis mõnes diskursuses esinevad väga harva. Näiteks siinses töös on korpusmaterjalile tuginedes vaadeldud määratlejatena esinevate demonstratiivadverbide ja demonstratiivpronoomenite referentsiaalseid erinevusi. Kvantitatiivse korpusanalüüsi tarbeks kogutud näitelause korral ei ole uurijal tavaliselt võimalik süveneda lausete laiemasse esinemiskonteksti või lausa tervesse diskursusesse. Seetõttu saab korpusmaterjali põhjal analüüsida referentsiaalseid omadusi n-ö lokaalsel tasandil, keskendudes uuritava konstruktsiooni semantilisele või grammatilisele ümbrusele lauses. Sellisel tasandil on kasulik võrrelda nt omavahel sarnaseid konstruktsioone, mille erinevused avalduvadki üldiselt semantilisel tasandil (konstruktsioonilise varieerumise uurimise kohta vt nt Klavan 2012; Arppe 2009; Silvennoinen 2018).

ii) Erinevate fraaside (või konstruktsioonide) kodeerimisega seotud protseduuri täpsus ja läbipaistvus osutub eriti tähtsaks neil juhtudel, mil vaadeldavatel fraasidel on väga sarnane tähendus ning neid kasutatakse referentsiaalselt lähedastes kontekstides. Doktoritöös oli lähema vaatluse all kaks sellist paari: i) kolmanda isiku pronoomen *ta* ning sellega lähedane nullviitamine (artikkel ZERO) ning ii) demonstratiivadverbist määratlejaga *vs.* demonstratiivpronoomenist määratlejaga definiitsed konstruktsioonid (artikkel DEMONSTRATIVES). Neid paarikuid eristab see, et esimese puhul on üks vaadeldavatest fraasidest n-ö nähtamatu, st ei avaldu lauses fonoloogiliselt (nullviitamine). Teise paari omapära peitub aga selles, et üks huvipakkuv konstruktsioon (dem-adv + NP) esines suulistes narratiivides üksikutel kordadel, samas kui teine konstruktsioon (dem-pron + NP) oli suhteliselt sage. Just neist erinevustest ongi tingitud nende uurimuste metodoloogilised lahknevused nii kodeerimisel kui ka seletavate tunnuste valikul.

## Kokkuvõte

Siinne doktoritöö on esimene süstemaatiline mitmeteguriline eesti keele viitevahendite süsteemi käsitus, mis vaatleb viitevahendite esinemist ja funktsioone läbivalt kogu diskursuses. Töö keskne fookus on sellel, miks ja millistes kontekstides erinevaid viitevahendeid diskursusesse valitakse ja kuidas need vahelduvad. Doktoritöö tulemused ei ole relevantssed vaid eesti keele süsteemi seletamiseks, vaid on laiendatavad ka teistele keeltele. Doktoritöö olulisim järeldus on, et eesti keele ja laiemalt mistahes keele viitevahendite süsteemi ja kasutust ei ole otstarbekas selgitada vaid viidatava referendi esilduvusele tuginedes. Viitevahendi valikut diskursuses mõjutab korraga mitu erinevat tegurit ning erinevad viitevahendid võivad olla seotud täiesti erinevate teguritega. Viitamist mõjutavate tegurite tervikkomplekt ei ole aga universaalne, vaid see on iga keele ja ka kõneleja puhul omanäoline.

## **PUBLICATIONS**

## CURRICULUM VITAE

Name: Helen Hint  
Date of birth: 22.05.1987  
Citizenship: Estonian  
Telephone: +372 5559 0644  
E-mail: helen.hint@ut.ee

### Education:

2017– University of Tartu, Institute of Education, Teacher Education (MA studies)  
2012–2021 University of Tartu, Institute of Estonian and General Linguistics, Department of General Linguistics (doctoral studies)  
2014–2015 University of Southern California, Department of Linguistics (visiting PhD student)  
2009–2012 University of Tartu, Institute of Estonian and General Linguistics, Department of General Linguistics (MA *cum laude*)  
2006–2009 University of Tartu, Institute of Estonian and General Linguistics, Department of General Linguistics (BA)  
2003–2006 Hugo Treffner Gymnasium  
1994–2003 Tartu Karlova Gymnasium

### Professional employment:

2020–2023 University of Tartu, Faculty of Arts and Humanities, Institute of Estonian and General Linguistics, Junior Research Fellow in Academic Writing  
2019– Tallinn University, School of Humanities, Lecturer of linguistics  
2017–2020 University of Tartu, Faculty of Social Sciences, Institute of Education, Junior Research Fellow in Educational Science  
2014 University of Tartu, Faculty of Philosophy, Institute of Estonian and General Linguistics, Junior Research Fellow  
2010–2016 University of Tartu, Faculty of Philosophy, Institute of Estonian and General Linguistics, Teaching assistant

### Publications:

Hint, Helen, Piia Taremaa, Maria Reile and Renate Pajusalu. 2021. Demonstratiivpronoomenid ja -adverbid määratlejatena: miks me oleme *siin ilmas, selles olukorras*? [Demonstrative pronouns and demonstrative adverbs as determiners in Estonian: why are we in “here world” in “this situation”?]. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics* (in press).

Hint, Helen and Anni Jürine. 2021. Protsessikeskse kirjutamise kasulikkusest gümnaasiumiõpilaste kirjaliku tekstiloome oskuse kui suhtluspädevuse ühe komponendi arendamisel [The usefulness of process-oriented writing in enhancing writing skills as a component of communicative competence

- among upper secondary school students]. *Eesti Haridusteaduste Ajakiri* = *Estonian Journal of Education* 9(1). 243–273.
- Taremaa, Piia, Helen Hint, Maria Reile and Renate Pajusalu. 2021. Constructional Variation in Estonian: Demonstrative Pronouns and Adverbs as Determiners in Noun Phrases. *Lingua* 254. 103030.
- Hint, Helen, Tiina Nahkola and Renate Pajusalu. 2020. Pronouns as referential devices in Estonian, Finnish, and Russian. *Journal of Pragmatics* 155. 43–63.
- Pajusalu, Renate, Helen Hint, Maria Reile and Piia Taremaa. 2020. Eesti keele viitevahendid kirjeldustes ja jutustuses [Reference in Estonian descriptions and narratives]. *Keel ja Kirjandus* 4. 313–335.
- Hint, Helen, Anni Jürine and Ilona Tragel. 2020. Suhtluspädevus [Communicative competence]. In Margus Pedaste (Ed.). *Üldpädevused gümnaasiumis*, 133–158. Tartu: University of Tartu Press.
- Pajusalu, Renate, Maria Reile, Helen Hint, Tiina Nahkola and Piia Taremaa. 2018. Relative clauses in spatial and narrative contexts in Estonian, Finnish, and Russian. *SKY Journal of Linguistics*, 31, 107–142.
- Hint, Helen, Tiina Nahkola and Renate Pajusalu. 2017. With or without articles? A comparison of article-like determiners in Estonian and Finnish. *Lähivõrdlusi. Lähivertailuja* 27. 65–106.
- 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.
- Hint, Helen, Maria Reile and Renate Pajusalu. 2013. Kontekst ja viitamine: argivestlused, legod ja narratiivid [Context and reference: everyday conversations, Lego blocks and narratives]. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics* 3. 161–183.
- Kirsipuu, Helen, Piret Soodla and Renate Pajusalu. 2012. Referentsiaalsed noomenifraasid laste narratiivides [Referential noun phrases in Estonian children's narratives]. *Eesti Rakenduslingvistika Ühingu aastaraamat* 8. 91–107.

## ELULOOKIRJELDUS

Nimi: Helen Hint  
Sünniaeg: 22.05.1987  
Kodaondsus: Eesti  
Telefon: +372 5559 0644  
E-post: helen.hint@ut.ee

### Haridustee:

2017– Tartu Ülikool, haridusteaduste instituut, õpetajaharidus (magistri-  
õpe)  
2012–2021 Tartu Ülikool, eesti ja üldkeeleteaduse instituut, üldkeeleteadus  
(doktoriõpe)  
2014–2015 Lõuna-California Ülikool (USA), keeleteaduse osakond (külalis-  
doktorant)  
2009–2012 Tartu Ülikool, eesti ja üldkeeleteaduse instituut, üldkeeleteadus  
(MA *cum laude*)  
2006–2009 Tartu Ülikool, eesti ja üldkeeleteaduse instituut, üldkeeleteadus (BA)  
2003–2006 Hugo Treffneri Gümnaasium  
1994–2003 Tartu Karlova Gümnaasium

### Töökogemus:

2020–2023 Tartu Ülikool, humanitaarteaduste ja kunstide valdkond, eesti ja  
üldkeeleteaduse instituut, akadeemilise väljendusoskuse noorem-  
teadur  
2019– Tallinna Ülikool, humanitaarteaduste instituut, lingvistika lektor  
2017–2020 Tartu Ülikool, sotsiaalteaduste valdkond, haridusteaduste instituut,  
haridusteaduste nooremteadur  
2014 Tartu Ülikool, filosoofiateaduskond, eesti ja üldkeeleteaduse  
instituut, nooremteadur  
2010–2016 Tartu Ülikool, filosoofiateaduskond, eesti ja üldkeeleteaduse insti-  
tuut, õppeassistent

### Publikatsioonid:

Hint, Helen, Piia Taremaa, Maria Reile ja Renate Pajusalu. 2021. Demonstra-  
tiivpronoomenid ja -adverbid määratlejatena: miks me oleme *siin ilmas*,  
*selles olukorras*? *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of*  
*Estonian and Finno-Ugric Linguistics*, (trükkis).  
Hint, Helen ja Anni Jürine. 2021. Protsessikeskse kirjutamise kasulikkusest  
gümnaasiumiõpilaste kirjaliku tekstiloome oskuse kui suhtluspädevuse ühe  
komponendi arendamisel. *Eesti Haridusteaduste Ajakiri = Estonian Journal*  
*of Education* 9(1), 243–273.  
Taremaa, Piia, Helen Hint, Maria Reile ja Renate Pajusalu. 2021. Construc-  
tional Variation in Estonian: Demonstrative Pronouns and Adverbs as Deter-  
miners in Noun Phrases. *Lingua* 254. 103030.



- Hint, Helen, Tiina Nahkola ja Renate Pajusalu. 2020. Pronouns as referential devices in Estonian, Finnish, and Russian. *Journal of Pragmatics* 155. 43–63.
- Pajusalu, Renate, Helen Hint, Maria Reile ja Piia Taremaa. 2020. Eesti keele viitevahendid kirjeldustes ja jutustuses [Reference in Estonian descriptions and narratives]. *Keel ja Kirjandus* 4. 313–335.
- Hint, Helen, Anni Jürine ja Ilona Tragel. 2020. Suhtluspädevus. In Margus Pedaste (Toim.). *Üldpädevused gümnaasiumis*, 133–158. Tartu: Tartu Ülikooli Kirjastus.
- Pajusalu, Renate, Maria Reile, Helen Hint, Tiina Nahkola ja Piia Taremaa, Piia. 2018. Relative clauses in spatial and narrative contexts in Estonian, Finnish, and Russian. *SKY Journal of Linguistics* 31. 107–142.
- Hint, Helen, Tiina Nahkola ja Renate Pajusalu. 2017. With or without articles? A comparison of article-like determiners in Estonian and Finnish. *Lähi-võrdlusi. Lähivertailuja* 27. 65–106.
- 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.
- Hint, Helen, Maria Reile ja Renate Pajusalu. 2013. Kontekst ja viitamine: argi-vestlused, legod ja narratiivid. *Eesti ja soome-ugri keeleteaduse ajakiri / Journal of Estonian and Finno-Ugric Linguistics* 3. 161–183.
- Kirsipuu, Helen, Piret Soodla ja Renate Pajusalu. 2012. Referentsiaalsed noomenifraasid laste narratiivides. *Eesti Rakenduslingvistika Ühingu aasta-raamat*, 8, 91–107.

## DISSERTATIONES LINGUISTICAE UNIVERSITATIS TARTUENSIS

1. **Anna Verschik.** Estonian Yiddish and its contacts with coteritorial languages. Tartu, 2000, 196 p.
2. **Silvi Tenjes.** Nonverbal means as regulators in communication: socio-cultural perspectives. Tartu, 2001, 214 p.
3. **Ilona Tragel.** Eesti keele tuumverbid. Tartu, 2003, 196 lk.
4. **Einar Meister.** Promoting Estonian speech technology: from resources to prototypes. Tartu, 2003, 217 p.
5. **Ene Vainik.** Lexical knowledge of emotions: the structure, variability and semantics of the Estonian emotion vocabulary. Tartu, 2004, 166 p.
6. **Heili Orav.** Isiksuseomaduste sõnavara semantika eesti keeles. Tartu, 2006, 175 lk.
7. **Larissa Degel.** Intellektuaalsfäär intellektuaalseid võimeid tähistavate sõnade kasutuse põhjal eesti ja vene keeles. Tartu, 2007, 225 lk.
8. **Meelis Mihkla.** Kõne ajalise struktuuri modelleerimine eestikeelsele tekst-kõne sünteesile. Modelling the temporal structure of speech for the Estonian text-to-speech synthesis. Tartu, 2007, 176 lk.
9. **Mari Uusküla.** Basic colour terms in Finno-Ugric and Slavonic languages: myths and facts. Tartu, 2008, 207 p.
10. **Petar Kehayov.** An Areal-Typological Perspective to Evidentiality: the Cases of the Balkan and Baltic Linguistic Areas. Tartu, 2008, 201 p.
11. **Ann Veismann.** Eesti keele kaas- ja mäarsõnade semantika võimalusi. Tartu, 2009, 145 lk.
12. **Erki Luuk.** The noun/verb and predicate/argument structures. Tartu, 2009, 99 p.
13. **Andriela Rääbis.** Eesti telefonivestluste sissejuhatuse struktuur ja suhtlusfunktsioonid. Tartu, 2009, 196 lk.
14. **Liivi Hollman.** Basic color terms in Estonian Sign Language. Tartu, 2010, 144 p.
15. **Jane Klavan.** Evidence in linguistics: corpus-linguistic and experimental methods for studying grammatical synonymy. Tartu, 2012, 285 p.
16. **Krista Mihkels.** Keel, keha ja kaardikepp: õpetaja algatatud parandussekventsides multimodaalne analüüs. Tartu, 2013, 242 lk.
17. **Sirli Parm.** Eesti keele ajasõnade omandamine. Tartu, 2013, 190 lk.
18. **Rene Altrov.** The Creation of the Estonian Emotional Speech Corpus and the Perception of Emotions. Tartu, 2014, 145 p.
19. **Jingyi Gao.** Basic Color Terms in Chinese: Studies after the Evolutionary Theory of Basic Color Terms. Tartu, 2014, 248 p.
20. **Diana Maisla.** Eesti keele mineviku ajavormid vene emakeelega üliõpilaste kasutuses. Tartu, 2014, 149 lk.

21. **Kersten Lehismets.** Suomen kielen väylää ilmaisevien adpositioiden *yli, läpi, kautta ja pitkin* kognitiivista semantiikkaa. Tartu, 2014, 200 lk.
22. **Ingrid Rummo.** A Case Study of the Communicative Abilities of a Subject with Mosaic Patau Syndrome. Tartu, 2015, 270 p.
23. **Liisi Piits.** Sagedamate inimest tähistavate sõnade kollokatsioonid eesti keeles. Tartu, 2015, 164 lk.
24. **Marri Amon.** Initial and final detachments in spoken Estonian: a study in the framework of Information Structuring. Tartu, 2015, 216 p.
25. **Miina Norvik.** Future time reference devices in Livonian in a Finnic context. Tartu, 2015, 228 p.
26. **Reeli Torn-Leesik.** An investigation of voice constructions in Estonian. Tartu, 2015, 240 p.
27. **Siiri Pärkson.** Dialoogist dialoogsüsteemini: partneri algatatud parandused. Tartu, 2016, 314 lk.
28. **Djuddah A. J. Leijen.** Advancing writing research: an investigation of the effects of web-based peer review on second language writing. Tartu, 2016, 172 p.
29. **Piia Taremaa.** Attention meets language: a corpus study on the expression of motion in Estonian. Tartu, 2017, 333 p.
30. **Liina Tammekänd.** Narratological analysis of Võru-Estonian bilingualism. Tartu, 2017, 217 p.
31. **Eva Ingerpuu-Rümmel.** Teachers and learners constructing meaning in the foreign language classrooms: A study of multimodal communication in Estonian and French classes. Tartu, 2018, 218 p.
32. **Kaidi Rätsep.** Colour terms in Turkish, Estonian and Russian: How many basic blue terms are there? Tartu, 2018, 181 p.
33. **Kirsi Laanesoo.** Polüfunktsionaalsed küsilauseid eesti argivestluses. Tartu, 2018, 176 lk.
34. **Maria Reile.** Estonian demonstratives in exophoric use: an experimental approach. Tartu, 2019, 240 p.
35. **Helen Türk.** Consonantal quantity systems in Estonian and Inari Saami. Tartu, 2019, 149 p.
36. **Andra Rumm.** Avatud küsimused ja nende vastused eesti suulises argivestluses. Tartu, 2019, 217 lk.
37. **Eleri Aedmaa.** Detecting Compositionality of Estonian Particle Verbs with Statistical and Linguistic Methods. Tartu, 2019, 271 p.
38. **Kristina Koppel.** Näitelausete korpuspõhine automaattuvastus eesti keele õppesõnastikele. Tartu, 2020, 249 lk.
39. **Ilze Tālberga.** On the equivalents of the Latvian verbal prefixes in Estonian. Tartu, 2020, 210 p.
40. **Roger M. A. Yallop.** The affect and effect of asynchronous written artefacts (cover letters, drafts, and feedback letters) within L2 English doctorate writing groups. Tartu, 2020, 312 p.
41. **Mariann Proos.** Meaning and usage of Estonian experience perception verbs. Tartu, 2021, 175 p.