

MARI AIGRO

In any case?  
Estonian spatial cases  
as argument markers





**MARI AIGRO**

In any case?  
Estonian spatial cases as argument markers



UNIVERSITY OF TARTU

Press

Institute of Estonian and General Linguistics, University of Tartu

Dissertation accepted for the defence of the degree of Doctor of Philosophy on October 13, 2022 by the Committee of the Institute of Estonian and General Linguistics, Faculty of Arts and Humanities, University of Tartu

Supervisors: Associate Prof. Virve-Anneli Vihman, PhD (University of Tartu)  
Dr Merilin Miljan, PhD (University of Tallinn)

Opponent: Prof. Dagmar Divjak, PhD (University of Birmingham, UK)

Commencement: December 9, 2022 at 11.00 at Ülikooli 18-204, Tartu.

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



European Union  
European Regional  
Development Fund



Investing  
in your future

ISSN 1406-5657 (print)  
ISBN 978-9916-27-090-5 (print)  
ISSN 2806-237X (pdf)  
ISBN 978-9916-27-091-2 (pdf)

Copyright: Mari Aigro, 2022

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

*To mum and dad.*

## ACKNOWLEDGMENTS

After I started my second undergraduate programme in Aberdeen, Scotland in 2013, this time with a clear goal of never leaving academia, I did not expect that I was going to be returning to Tartu for my PhD. While I did love Tartu, having spent some time here during my first (unsuccessful) undergraduate programme, returning here seemed unlikely, because ideas about how to be a great academic had already started to settle in my mind. The path forward seemed to be filled with a number of postdoc positions in various countries, never spending more than two years at a single university, never settling down, always looking for new projects with new research teams. The future seemed to roll out like a patchwork blanket and I thought I better go and get myself a really nice set of suitcases that would withstand the frequent moving.

However, the lecturers at my first real alma mater, University of Aberdeen, did not appear to lead such lives. Instead, their international academic collaborations seemed to happily co-exist side by side with their values about their community, the city in which they were building their lives, their home. It made me think about my home, convincing me I need to return there. I could not be happier about having made this decision.

When I arrived in Tartu, I had two goals — getting my PhD and finding that same sort of community of people. I feel privileged that both wishes were granted and I realise I need to thank a long list of people for this. However luring it is to become fixated on one's career in the modern society, in the end all careers are really only about the people involved in them and around them.

Even though Aberdeen was a cold windy granite city, I have nothing but warm memories of its University. Lecturers there set the tone for much of what I have done after leaving Aberdeen. I am not going to forget Robert Millar, Elspeth Edelstein and Will Barras, who not only gave me a compassionate, people-oriented perspective on how to study language, but who also showed me what kind of a researcher-lecturer I want to become. I am lucky if I ever match their warmth, consideration, support and wit. I especially will not forget Robert, my first mentor, and the many linguistics-related and non-related chats in his office. Your advice have always stuck with me.

My time in Cambridge was full of personal and work-related growth, which was sometimes challenging to keep at a sustainable pace. I am grateful to my supervisor, Jim Blevins, for having guided me through it, even when I felt like a fish out of water at times. I will also never forget our kitchen evenings with Mar and Beatrice, who injected peace and lightness in my life on a daily basis.

This brings me to Tartu and, first and foremost, to Virve, my PhD supervisor. Virve has worn many hats during my graduate studies. She has been my main guide in the academic world, my mentor, my teacher in linguistics, my therapist, my friend. I am not sure a supervisor can be expected to be all that, which is why I cannot help but feel I lucked out with her. Whatever shape my career will

take, Virve helped build it. She was the one going back and forth, wheeling in the bricks, mixing the mortar, explaining the blueprints and brick building 101 to me. I am forever grateful for that. At times this meant connecting me to a network of researchers, handing me the tools and showing how they worked. Other times it just meant keeping me going, listening to the the anxiety and desperation I occasionally needed to pour out. Yet other times it meant showing me how to be a mum and a researcher, or that it is possible to be both. I feel privileged to have had someone believe in me and be there for me in such an unconditional way throughout these years.

I also feel grateful for the support of many other people in the institute who have created a supportive and intellectually rich work environment. This includes my writing group with Piia, Maria, Helen and others, Enel and other members of the kitchen crew, fellow PhD students as well as Liina, Kaidi and others. I am grateful to my other supervisor, Merilin, for her guidance, feedback and her work on the thesis. I am also grateful to Piia for her thorough feedback on my thesis, as well as to the external reviewers, Dr. Heete Sahkai and Prof. Dagmar Divjak, for their much needed comments. In addition, Prof. Helen de Hoop and Sander Lestrade gave me very helpful feedback on my work during my time in Nijmegen. Finally, my gratitude goes to colleagues in the overabundance/defectivity project, the teenager language project and the concreteness/abstractness project for providing me with an unusually stimulating work environment.

I am not taking it for granted that several people in the institute have grown into close personal friends. Anton and Mariann, this especially means you. Your friendship, perspective, humor and wit were there when I needed them the most, oftentimes with cake, helping me keep my priorities straight and making me laugh on a daily basis. You made it easy and light and that is no small feat.

I was also lucky to have been randomly given a work space in room 430, the best room in the building, the inhabitants of which form a cohesive, supportive community. Thank you for your encouragement, friendship and humor, Triin and especially Maarja-Liisa, who effectively became my third supervisor and a teacher of statistics. I appreciate you guys.

Zooming out, the rest of the community revolving around me is made of friends and family who stuck with me through thick (PhD) and thin (the rest). This means Monika, Eeva, Merli, Anu, Mihkel, Juudit, Berit, Jaan, as well as Kairiin and Merili, all of whom have my gratitude and admiration. This also includes the impressive women in my book club, to whom I am grateful for making me read when all my tired brain wanted to do was rewatch *The Office*.

My family, too, are a special group of people of whom I am very proud. Mum and dad to whom this thesis is dedicated — I cannot really find the words to describe the ways in which you have helped me. You have taught me to value knowledge, learning and the community of people around me, given me a love of literature and poetry as well as an active lifestyle. I have received nothing but

support and love from you throughout this journey and that is a privilege for which I am grateful. In addition, my gratitude goes to Indrek and Andres, my two older brothers to whom I have always looked up and who never fail to make me laugh. I appreciate the way you and your families have been there for me. This also goes for my supportive in-laws — thank you for being there and for taking my dog on long walks when I could not.

Finally, the spotlight narrows down to just a few people, my best people (and my best dog), the loves of my life — Enno, Jüri and Maia. Enno — it is a privilege that you have silently and patiently endured the frantic time period of writing up this thesis and various burnouts, and that you have answered my anxiety with nothing but calm smiles and helpful life advice. It is an even greater privilege that you have been my best friend for 11 years, my reference point. Thank you for your intelligence, your grounding perspective on life, your witty humor and your passion, all of which inspire me every day.

In a world fixated on economy and growth, doing linguistic research for a living feels like an act of heresy. However, even though research is a rewarding and interesting way to earn a living, even the best job in the world would make for a poor experience if it was laced with indifferent, self-centered people. I am grateful to the special community around me for being exactly the way it is. Because in the end, all experiences are shaped by people and it is only people that matter.

“When one stops talking about switches and begins to talk about potentiometers, one does not necessarily cease talking about electrical systems.”

Bolinger (1961, 10)

“Oh my god I am loving on you!”

Jonathan van Ness,  
co-host of *Queer Eye*

# CONTENTS

<b>ABBREVIATIONS</b>	<b>14</b>
<b>1 INTRODUCTION</b>	<b>16</b>
1.1 Outline of the thesis . . . . .	19
<b>2 BACKGROUND</b>	<b>21</b>
2.1 Terminology . . . . .	21
2.1.1 Arguments and adjuncts . . . . .	22
2.1.2 Complements and modifiers . . . . .	23
2.1.3 Valency and transitivity . . . . .	23
2.1.4 Argument structures or argument constructions? . . . . .	24
2.1.5 Subjects and objects . . . . .	25
2.1.6 Productivity of argument structures . . . . .	26
2.1.7 Argument status . . . . .	26
2.2 Argumenthood indicators . . . . .	27
2.2.1 Criteria accommodating all argument structures . . . . .	28
2.2.2 Highly problematic criteria . . . . .	31
2.2.3 Criteria describing the variation between argument structures . . . . .	39
2.2.4 Summary . . . . .	42
2.3 Cases as a dynamic system . . . . .	43
2.3.1 The origin of semantic cases . . . . .	44
2.3.2 From semantic cases to structural cases . . . . .	46
2.3.3 Economy in speech . . . . .	49
2.3.4 Summary . . . . .	50
2.4 Gradience in grammar . . . . .	51
2.4.1 Role-based approaches to argumenthood . . . . .	53
2.4.2 Gradient approaches to argumenthood . . . . .	55
2.4.3 Summary . . . . .	60
<b>3 AN OVERVIEW OF ESTONIAN</b>	<b>61</b>
3.1 Estonian nominal inflection . . . . .	61
3.2 Spatial cases . . . . .	63
3.2.1 Morphology . . . . .	64
3.2.2 Functions . . . . .	65
3.3 Grammatical relations . . . . .	69
3.3.1 Differential subject marking (DSM) . . . . .	69
3.3.2 Differential object marking (DOM) . . . . .	70
3.4 Semantic use of structural cases . . . . .	72
3.5 Structural use of semantic cases . . . . .	73
3.6 The current view . . . . .	77
3.6.1 Issues with the current view . . . . .	78

3.7	This thesis . . . . .	79
3.7.1	Verbs . . . . .	80
3.7.2	Cases . . . . .	81
3.7.3	Argument status . . . . .	81
3.7.4	Subject matter . . . . .	82
3.7.5	Questions and hypotheses . . . . .	82
<b>4</b>	<b>VERBS AND ARGUMENT STRUCTURES</b>	<b>84</b>
4.1	Data . . . . .	85
4.1.1	The corpus . . . . .	85
4.1.2	Detecting case selection in a corpus . . . . .	85
4.1.3	Oblique argument data . . . . .	86
4.1.4	Canonical argument data . . . . .	90
4.2	Study 1: The lexical transitivity of verbs . . . . .	91
4.2.1	Method . . . . .	91
4.2.2	Results . . . . .	97
4.3	Study 2: Bleached use of case . . . . .	106
4.3.1	Method . . . . .	106
4.3.2	Results . . . . .	108
4.4	Study 3: Usage patterns of argument structures . . . . .	110
4.4.1	Method . . . . .	110
4.4.2	Results of mixed logistic regression . . . . .	112
4.4.3	Structure-based analysis . . . . .	115
4.5	Summary . . . . .	123
4.5.1	Discussion . . . . .	125
<b>5</b>	<b>CASES AND ARGUMENT STRUCTURES</b>	<b>127</b>
5.1	Introduction . . . . .	127
5.1.1	Functional profiles . . . . .	128
5.1.2	Lexical selection patterns . . . . .	130
5.1.3	Syntactic productivity . . . . .	131
5.1.4	Research questions . . . . .	132
5.2	Study 1: Functional profiles of cases . . . . .	133
5.2.1	Data and method . . . . .	133
5.2.2	Results . . . . .	138
5.2.3	Summary . . . . .	140
5.3	Study 2: Lexical selection patterns of cases . . . . .	140
5.3.1	Data . . . . .	140
5.3.2	Method . . . . .	141
5.3.3	Results: all case tokens . . . . .	142
5.3.4	Results: spatial cases in governed position . . . . .	144
5.3.5	Summary . . . . .	144
5.4	Study 3: Syntactic productivity . . . . .	145
5.4.1	Data and method . . . . .	145
5.4.2	Results: syntactic productivity in general . . . . .	147

5.4.3	Summary . . . . .	149
5.5	Summary . . . . .	150
5.5.1	From spatial meaning to structural use . . . . .	151
5.5.2	From comitative meaning to structural use . . . . .	155
<b>6</b>	<b>ARGUMENT STATUS ON A GRADIENT</b>	<b>158</b>
6.1	Arguments in experimental approaches . . . . .	159
6.1.1	Sentence processing . . . . .	159
6.1.2	Measuring argument status . . . . .	161
6.1.3	Semi-arguments . . . . .	164
6.2	Argument judgment task — aims and research questions . . . . .	165
6.2.1	Oblique arguments . . . . .	166
6.2.2	Participant locations . . . . .	167
6.2.3	Research questions . . . . .	168
6.3	Experiment method . . . . .	168
6.3.1	Participants . . . . .	168
6.3.2	Material . . . . .	168
6.3.3	Trials . . . . .	170
6.3.4	Procedure . . . . .	173
6.4	Results . . . . .	174
6.4.1	Argument status of elements marked in spatial cases . . . . .	175
6.5	Summary . . . . .	178
6.5.1	Gradience in argumenthood . . . . .	179
6.5.2	Factors affecting argument status . . . . .	180
<b>7</b>	<b>DISCUSSION</b>	<b>181</b>
7.1	The diachronic perspective . . . . .	181
7.2	The synchronic perspective . . . . .	183
7.2.1	Verbs and argument structures . . . . .	183
7.2.2	Argument status . . . . .	185
7.2.3	Argument structures as complementary patterns . . . . .	186
7.2.4	Canonicity and prototypicality in gradient argument status . . . . .	188
7.3	The origin of obliques . . . . .	189
<b>8</b>	<b>SUMMARY IN ESTONIAN</b>	<b>192</b>
8.1	Sissejuhatus . . . . .	192
8.1.1	Argumendistruktuuride süsteem . . . . .	192
8.1.2	Eesti keel kui haruldaset kasulik allikas . . . . .	193
8.2	Käesolev töö – eesmärgid ja uurimisküsimused . . . . .	194
8.3	Verbid argumendistruktuurides . . . . .	196
8.3.1	Meetod . . . . .	196
8.3.2	Tulemused . . . . .	198
8.4	Käänded argumendistruktuurides . . . . .	199
8.4.1	Meetod . . . . .	200
8.4.2	Tulemused . . . . .	201

8.5	Argumendistaatus argumendistruktuurides . . . . .	202
8.5.1	Meetod ja küsimused . . . . .	203
8.5.2	Tulemused . . . . .	204
8.6	Kokkuvõte . . . . .	205
<b>REFERENCES</b>		<b>208</b>
<b>CURRICULUM VITAE</b>		<b>220</b>
<b>ELULOOKIRJELDUS</b>		<b>221</b>

## ABBREVIATIONS

### Abbreviations in glosses<sup>1</sup>

1, 2, 3	person	INE	inessive case
ABL	ablative case	INF1	da-infinitive
ACC	accusative case	NEG	negative
ADE	adessive case	NOM	nominative case
ALL	allative case	PAR	partitive case
AUX	auxiliary verb	PL	plural
CNG	connegative	POS	possessive
COM	comitative case	POST	postposition
COP	copula	PPP	passive past participle
CVB	converb	PPT	past participle
DAT	dative case	PRS	present tense
ELA	elative case	PST	past simple tense
EVID	evidential	REFL	reflexive
GEN	genitive case	SG	singular
ILL	illative case	TERM	terminative case
IMP	imperative		
IMPRS	impersonal		

---

<sup>1</sup> This thesis segments glosses slightly differently from the Leipzig Glossing Rules (<https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf>). Namely, the rules maintain that hyphens should be used to separate the reference form (the nominative form for nominals) from a suffix. In this thesis, however, the hyphen is also used with non-nominative reference forms, where the affix is nevertheless agglutinative. This is because most Estonian case suffixes are added to the genitive rather than the nominative form. Genitive case morphology, however, is not agglutinative but fusional, meaning the stem is not separable from the suffix, but the two form a new, single morpheme. Hyphens are not used in tokens with more profound stem changes.

## Other abbreviations

K-state	Kimian state
D-state	Davidsonian state
DCM	Differential case marking
DSM	Differential subject marking
DOM	Differential object marking
n1	<i>hapax legomena</i> count of a morpheme
Nc	token count
NP	Noun phrase
P	Potential variance
pmw	per million words
RQ	research question
V	Realised variance (type count)

# 1 INTRODUCTION

This thesis addresses a well-known issue in the syntax-semantics interface — the argument-adjunct dichotomy. More precisely, it is concerned with argument structures where verbs select semantic cases for marking their complements. Such complements are sometimes regarded as non-canonically marked arguments, because semantic cases are commonly regarded as markers of adjuncts rather than arguments. For instance, in (1a), the Estonian elative case (‘out of’) marks physical space while in (1b) it marks the complement of the verb *sõltuma* ‘depend on’:

- (1) a. Enno kōndi-s bussijaama-st koju  
Enno.NOM walk-PST.3SG bus station-ELA home.ILL  
‘Enno walked home from the bus station’  
b. Üritus sõltu-b ilma-st  
event.NOM depend-PRS.3SG weather-ELA  
‘The event depends on the weather’

Morphological case is a widely used way of expressing verbal argument structure in language. It is also commonly assumed that some cases are better described as ‘structural’ (sometimes also called ‘syntactic’ or ‘grammatical’) while others are ‘semantic’. While structural cases are expected to mark core arguments (subjects and objects), semantic cases are viewed as prototypical adverbial markers (e.g. expressing spatial information, e.g. *koju* in (1a)). Hence, arguments marked in semantic cases (1b) complicate this seemingly simple division of labour between case types. Henceforth they will be called oblique arguments in this thesis.

Another way to approach case systems is to outline case functionality in the system of a given language and first see whether the two labels (‘structural’ and ‘semantic’) are even helpful in describing the system. Nichols (1983) posits that while they may apply to some languages, others have a case system with functionally overlapping cases, in which case using a binary labelling system would be misleading. Such a language has a richer inventory of linguistic devices for marking verbal arguments.

Estonian is a language in which morphological cases may be expected to have functionally mixed use. It is a morphologically rich language, where the nominal paradigm includes 14 cases. Three of them — partitive, genitive and nominative — are regarded as structural, marking canonical subjects and objects. These three cases alternate both in subjects and objects, meaning Estonian has both differential subject and object marking (DOM and DSM). The rest, i.e. 11 cases, are regarded as semantic. They include six spatial cases, which form a cohesive semantic system. Three mark internal space (illative ‘into’, inessive ‘in’ and elative ‘out of’) and three mark external space (allative ‘onto’, adessive ‘on’ and ablative ‘from on’). Both groups include three directions: lative (illative, allative), locative (inessive, adessive) and separative (elative, ablative).

Many of the Estonian semantic cases have been noted to be multifunctional in that they mark adverbials as well as occur in structural functions, e.g. elative in (1a)–(1b). However, not much is known about their structural function in terms of the extent of the phenomenon, how semantic case selection is restricted with regard to verb lexical semantics, how it affects verb usage, how semantic cases compare to each other in terms of the proportion in which they occur in structural function and how various semantic cases relate to argument status when used structurally.

The structural cases of Estonian, i.e. nominative, genitive and partitive, also offer little functional clarity. Not only do they alternate based on various semantic, pragmatic and syntactic factors when marking subjects and objects, but they also mark various phrases which can be described as showing both object and adverbial properties. All in all, the functionality of Estonian cases does not appear to map neatly on the labels ‘semantic’ and ‘structural’, as many can be seen in a wide range of different functions.

This thesis is interested in the structural use of semantic case. More precisely, it is interested in a particular type of structural use, expressed as a verbal argument structure. Syntactically, an argument structure is a syntactic co-occurrence pattern in which a verb selects a particular case for marking one or more select constituents that express the participants of the event or state denoted by the verb. Semantically, an argument structure is expected to incorporate cases which are used with bleached, non-coherent semantics, as their main function is to identify event participants.

A language with a case system can have an inventory of argument structures that vary in terms of selected case and number of participants. Marantz (2013, 152–153) summarises two puzzles pertaining to verbs and structures, as shown below. These quotes also illustrate why a description might benefit from a more cohesive concept of a structure, rather than merely referring to instances of case selection:

[...] the connection between syntactic structure — in the shallow sense of word order and morphological marking — and meaning, while varied within a language, also transcends individual verbs, such that we can productively talk about verbs choosing frames relating form and meaning from a small finite set, rather than about each verb constructing a particular form/meaning connection.

[...] the flexibility of verbs to appear within the various set of frames relating form and meaning has defied these efforts to regulate apparent alternations in argument structure through the classification of verbs.

These two lines of research have shifted discussion away from verb classes and verb-centered argument structures to the detailed analysis of the way that structure is used to convey meaning in languages.

(Marantz, 2013, 152–153)

When an argument structure is defined by cases and the number of arguments, it follows that a highly varied set of verbs must make do with a limited, finite set of argument structures for marking their state-of-affairs (event or state) participants. Hence, the structures must be highly generalised. Nevertheless, structures must also bear some degree of meaning themselves as both verbs and argument structures may be viewed as contributing to the meaning of the overall expression — verbs through their root semantics and structures through their structured meaning representation. Structures therefore contribute highly generalised, abstract meaning to the expression.

Structural cases and the argument structures incorporating them (henceforth ‘canonical argument structures’) have traditionally enjoyed a great deal more spotlight than semantic cases in structural functions. This has led to significant gaps in the knowledge about how argument structures vary in terms of assignment to verbs and usage, the degree to which the semantic cases in a language can participate in this phenomenon, as well as the type of argument status they mark. In addition, many studies that do focus on the structural function of semantic cases, do so by regarding it as entirely independent from their other functions (see section 2.2.3). However, case functions may be expected to be linked to one another, because structural functions diachronically arise from the extension of semantic functions, leading to grammaticalisation. It is not clear how the overall functional properties and the degree of grammaticalisation of semantic cases affects the way speakers use argument structures with these cases.

In addition to variation in terms of verbs and case properties, argument structures are also hypothesised to vary in terms of the nature of the underlying argument status, or in other words, the nature of the connection between verbs and various other constituents. This connection is what the concept of argument structures effectively models: a link connecting the verb to a few other concepts. A long held view by many authors is that structural cases indicate a stronger or a different connection between verbs and the nominal phrases they mark, compared to that between verbs and constituents in semantic cases. Others disagree, arguing for analogous argument status for arguments marked with both types of cases (Nichols, 1983). Even though we have the behavioural data based empirical paradigms which are necessary for comparing argument status quality in various structures, no studies have taken such an approach.

Overall, this thesis undertakes a comprehensive investigation of Estonian argument structures incorporating spatial cases (‘oblique argument structures’), using both corpus and behavioural data. Estonian spatial cases, which make up six of its 11 semantic cases, have been noted to be highly multifunctional, incorporating both semantic and structural functions. By focusing on a single semantic domain (i.e. that of spatial relations), the thesis will be able to compare and outline these cases in a more thorough and meaningful way. The thesis has three broad objectives:

1. to investigate the distinctions between verbs in oblique and canonical argument structures with regard to lexical semantics and the usage patterns of verbs;
2. to link the functions of individual cases to the system of argument structures and show how changes in case meaning give rise to new structures;
3. to compare various structures in terms of argument status, using them to demonstrate the gradient nature of the connection between verbs and select few constituents.

Understanding oblique arguments is crucial for understanding why arguments are linguistically marked in the way they are and how existing linguistic devices take this function on. Argument structures form a system, the internal dynamics and patterns of which are not thoroughly understood as the focus of research in this domain has been to describe arguments in canonical argument structures and their connection to verbs, rather than the more peripheral structures. However, it is precisely the periphery which might provide meaningful insight into the phenomenon as a whole.

## 1.1 Outline of the thesis

Chapter 2 presents the background for the three research objectives listed above. First, it discusses various properties which have been assigned to canonical arguments and argument structures in linguistic literature. After a critical evaluation of a range of properties (or criteria), it highlights the features which may be assumed to distinguish between oblique and canonical arguments. Second, it elaborates on the grammaticalisation of cases and, in particular, the pathway between semantic and structural cases, showing that the two classes are parts of a scale rather than discrete categories. Third, it discusses the notion of argument status and the history of research viewing it as a gradient phenomenon.

Chapter 3 gives an overview of canonical and oblique arguments in Estonian, discussing the differential marking of subjects and objects as well as the functions of six Estonian spatial cases which are investigated in this thesis. Section 3.7 presents the outline of the corpus-based and experimental methodology used in this thesis as well as the broad research questions pursued.

Chapter 4 investigates the lexical semantic distinction between verbs with oblique and canonical argument structures, as well as the distinction between individual oblique structures. It presents the results of three corpus studies, based on different variables describing the same sample, which quantitatively addresses both the lexical semantics and the usage patterns of verbs in canonical and oblique argument structures.

Chapter 5 uses synchronic means to investigate the degree of grammaticalisation of the six Estonian spatial cases. This is offered as explanation for the ways and the degree to which spatial cases are used in structural functions, i.e. many of

the distinctions found in Chapter 4. It outlines three distinct corpus studies based on three different datasets, relating the characteristics of individual cases to their behaviour in argument marking functions.

Chapter 6 presents the behavioural data from judgment task based experiment, comparing the argument status of various types of arguments and adverbials.

In any case, the overall implications of the results of this thesis are discussed in chapter 7.

## 2 BACKGROUND

This chapter lays out the issues related to the three objectives of this thesis, as outlined in chapter 1. First, it maps the relevant terminology as used in this thesis (section 2.1). Next, it discusses the background for the three lines of investigation in this thesis.

The first line of investigation compares oblique and canonical argument structures, providing a general outline of the dynamics guiding the assignment of these two types of structures to verbs in Estonian. However, there are no clear guidelines in terms of what might constitute an informative basis for this comparison. A productive starting point for this is the list of argument criteria, which have been proposed to definitively identify arguments and distinguish them from adjuncts. Such criteria essentially serve the aim of highlighting the essence of argument structure and outline its many characteristic syntactic, morphological and semantic patterns. While some criteria classify both oblique and canonical structures as marking core arguments, others highlight a distinction between the two. The latter are of central interest in terms of the comparison presented in chapter 4. Hence, in order to identify potential distinct patterns of oblique and canonical structures, section 2.2 analyses a list of traditional argument criteria, highlighting their relevance to this thesis.

The second objective of this thesis is to link the evolution of individual cases to the system of argument structures. Section 2.3 discusses the grammaticalisation of morphological cases, focusing on the attested pathways between semantic and structural cases. It observes that instead of discrete categories of semantic and structural cases, a language may have a rather multifunctional set of cases, the grammaticalisation degrees of which appear to vary on a gradient.

The third objective of this thesis is to assess the quality of argument connection in various argument structures. As oblique arguments have been suggested to represent gradience in argumenthood, section 2.4 discusses various approaches to argumenthood, focusing on theories describing argument status as an inherently gradient phenomenon.

### 2.1 Terminology

This thesis takes a lexicalist view on verbs, which regards them as complex concepts that entail some degree of argument information and are organised by various rules and patterns of grammar. This means that verbs are viewed as lexically making reference to arguments, i.e. external participants, without which it would not be possible to conceive of the events the verbs express. This means the study describes argument structure phenomena by means of lexicalist terminology, which one may find in various projectionist argument structure approaches. For instance, this broadly includes studies using semantic roles or Proto-roles (Dowty, 1991; Aissen, 1999; Primus, 1999; Ackerman and Moore, 1999; Davis

and Koenig, 2000; Foley and Van Valin, Robert D., 1984; Van Valin, Robert D., 1990; Van Valin and LaPolla, 1997), as well as studies based on predicate decomposition (Jackendoff, 1976, 1990; Lakoff, 1970; Van Valin and LaPolla, 1997; Croft, 1998; Levin and Rappaport-Hovav, 2005).

However, the theoretical underpinnings of the present thesis largely revolve around terminology. This means that the thesis largely aims to conduct its investigations from an a-theoretical point of view, drawing conclusions from empirical data on language phenomena, which any linguistic framework would need to be able to explain.

### 2.1.1 Arguments and adjuncts

The present thesis uses the term ‘argument’ to refer to a lexico-semantic conceptual entity, which a lexicalist view sees as being entailed in the meaning of the verb. For instance, the meaning of ‘eat’ implies that there exists an entity that eats and an entity being eaten. These participants are commonly viewed as the arguments of that verb. Arguments contrast to other types of information, such as the location, time and manner of eating, which are assumed to be not as conceptually linked to the verb (Ackema, 2015). These types of information are therefore referred to as adjuncts. The two categories are exemplified in (2):

- (2) [Mari] [ate] [a burger] [in the garden]  
argument verb argument adjunct

The distinction between arguments and adjuncts has been an active topic of research since Tesnière (1959, translation 2015), who viewed arguments as the ‘main actors’ of the play and adjuncts as the ‘stage props’ (Tesnière, 2015). However, the notion of arguments can broadly be traced back to the late 19th century, when Frege (1891, via Williams, 2015, 80) proposed that without certain elements predicates are incomplete or unsaturated. As an analogue to formulas which include arguments, predicates, too, are viewed as structures with a fixed number of arguments.

Some approaches differentiate between syntactic and semantic arguments (linguistic arguments and event participants) (Barbu and Toivonen, 2016). The two groups differ in that semantic arguments may remain implicit in the clause while syntactic arguments may not (Dowty, 2000; Croft, 2001; Barbu and Toivonen, 2016). They are similar, however, in that both are entailed in the concept of the state-of-affairs expressed by the verb. Section 2.2.2 further discusses the notion of syntactic obligatoriness and highlights a range of issues related to it.

This thesis uses the term ‘argument’ to refer to both canonically marked arguments bearing structural cases and non-canonically marked arguments bearing semantic cases. This is uncontroversial, because most agree that oblique arguments constitute semantic arguments, even when they are regarded as fundamentally distinct from core (syntactic) arguments. Non-canonically marked arguments

are sometimes called oblique arguments or quasi-arguments (Schütze, 1995; Van Valin and LaPolla, 1997; Blake, 2001), as well as quirky arguments (Zaenen and Maling, 1984; Jónsson, 2003).

### **2.1.2 Complements and modifiers**

Complements and modifiers are somewhat but not entirely parallel concepts to arguments and adjuncts. They are used as structural rather than semantic notions. Hence, a complement is an argument filling the ‘slot’ introduced by the head of the phrase, e.g. in (2), ‘a burger’ is a complement of the verb ‘eat’. Similarly, ‘the garden’ is a complement of the preposition ‘in’. While core subjects (‘Mary’ in (2)) are regarded as verbal arguments, they are not categorised as complements. This is because on a clause level they are regarded as being on the same hierarchical level as the verb, rather than the dependant of the verb (Ackema, 2015). Finally, ‘in the garden’ is a modifier of the verb in (2).

### **2.1.3 Valency and transitivity**

Two other terms are inherently linked to argument structure. Valency is a term used for describing the number of arguments a verb takes. Originating from chemistry, it views verbs as analogous to atoms, the valency of which determines their capacity for combining with other atoms (Matthews, 2008). Approaches differentiating between syntactic and semantic arguments also differentiate between the syntactic valency of verbs and the semantic valency of predicates (Ackema, 2015). However, this thesis advocates for a single type of argument status, where patterns of syntactic obligatoriness in a particular language ought to be explained by means of other linguistic phenomena not inherently related to argument status (see section 2.2.2). Hence, only one concept of valency is used. Verbs taking one argument (‘laugh’) are therefore one-place verbs, while two arguments lead to two-place verbs (‘eat’) and the rare instance of three arguments result in three-place verbs (‘give’). These labels are used to refer to verbs in active mood. In the end, valency is a feature of a particular expression of mood and therefore subject to various valency-changing processes, e.g. passivisation.

Another term connected to argument structure is transitivity. In the literature it is used in two rather distinct meanings. First, it is used as an analogue to valency, describing the number of participants. One-place verbs are intransitive, two-place verbs transitive and three-place verbs ditransitive. Some view valency as referring to semantic arguments and transitivity to syntactic arguments (Ackema, 2015, 248).

However, transitivity has come to be used in a rather different meaning over the past decades. Hopper and Thompson (1980) used the label to describe a complex semantic-pragmatic feature of clauses, which varies in a gradient manner. The transitivity value of a clause depends on a number of variables, some

of which are lexical-semantic characteristics of verbs (kinesis, punctuality, number of participants), while others describe subjects (agency), objects (affectedness and individuation), or the entire clause (e.g. negation and aspect) (Hopper and Thompson, 1980). Each of these factors contributes to the overall transitivity of the clause. For instance, clauses with volitional subjects ('Mari ate chocolate.') are more transitive than clauses with non-volitional subjects ('Mari forgot my name.').

Importantly for the present thesis, the concept may also be used to describe the lexical-semantic transitivity of verbs. For instance, verbs describing dynamic, telic events ('fix') can be viewed as more transitive than verbs describing non-dynamic, atelic states-of-affairs ('love'). In order to avoid confusion, the present dissertation uses 'one-place verb' and 'two-place verb' to refer to the number of arguments. Transitivity, however, is used to refer to the lexical-semantic property which broadly expresses the degree to which subjects affect the event described by the verb, and to which objects are affected by it.

### **2.1.4 Argument structures or argument constructions?**

Argument structures (sometimes called case frames) are viewed as morphosyntactic devices, used for expressing the participants of the event described by the verb. While this definition may render a rather broad set of structures, the present thesis only focuses on argument structures which entail three particular characteristics. First, they entail a particular number of arguments. Metaphorically speaking, one may regard a structure as something that is made up of lexical slots: one slot for the verb and the rest for its arguments. Second, a structure defines case morphology for each argument. Third, the cases used in the structure have bleached semantics. For instance, when marking the argument of the Estonian verb *sõltuma* 'depend on', elative ('out of') case is bleached from its spatial meaning, unlike when it is used to mark the Source location of the verb *lahkuma* 'leave'. Importantly, the present thesis does not view bleached semantics as a prerequisite for argument status. It postulates no hypotheses about the nature of the argument structure of *lahkuma* 'leave'. It does, however, exclude such structures from its research domain as it focuses on argument structures with bleached semantics. This is what is essentially meant by 'structural use of semantic case'.

There is some controversy with using the term 'argument structure' in plural in a lexicalist approach. In much of the linguistic literature on these approaches, argument structure is used as a mass noun, referring to a single phenomenon with varying parts. However, as the quote of Marantz in chapter 1 illustrates, recent trends require that there exists a term which one can use to highlight the distinctions between the ways in which various types of argument structures are used to express meaning.

The plural form of argument structures would pose no difficulty in another theoretical paradigm, i.e. in constructionist approaches, e.g. in Construction Gram-

mar (CxG). As opposed to the basic lexicalist approach, CxG does not assume that argument information is contained in the verb's lexical entry. Instead, its lexical entry is viewed as encoding nothing but the idiosyncratic meaning of the verb, i.e. its root (Goldberg, 1995, 28). The root becomes integrated into constructions, which are separately learned form-function pairings with meanings of their own. It is therefore constructions, not verbs, that encode argument structure and grammatical relation information, although some verbs do 'inherently profile' participant information in their lexical specifications (Goldberg, 1995, 225).

Even though the idea of structure meaning (outside of lexical meaning) is more inherently natural to constructionist than lexical approaches, it poses no inherent difficulty on the level it is discussed here. Many of the research questions asked in this thesis address the structural level of clauses, e.g. the meaning and usage (productivity) of entire structures. Findings are relevant to both lexicalist and constructionist approaches, because lexicalist approaches, too, must be able to account for structural phenomena, in order to retain their status as empirically sound theories.

Hence, the present thesis views Estonian as having a number of 'argument structures'. On one hand this highlights the structural and semantic similarities between various cases being used structurally, all the while enabling this thesis to investigate their distinctions. Hence, all structures are similar in that they entail argument slots and predetermined (bleached) morphology, but they vary in terms of the number of argument slots and particular morphological cases.

Viewing the number of arguments as a factor by which argument structures vary complicates matters, because arguments may remain implicit. This thesis does not differentiate between 'John ate.' and 'John ate chocolate.' in terms of argument structure. Instead, it regards 'eat' as a two-place verb encoding two arguments, regardless of what occurs in any particular clause containing 'eat' (see justification for this in section 2.2.2). However, a verb may nevertheless occur in several different argument structures, which are distinct in terms of case. For instance, the Estonian verb *uskuma* 'believe' may occur with a partitive complement ('believe something') or an illative complement ('believe in something'), the latter rendering a different sense of that verb.

### 2.1.5 Subjects and objects

Many languages have a predominant one-place and two-place argument structure, used by a wide range verbs and involved in various syntactic operations (Nichols, 1983). In languages with a nominative-accusative alignment, nominative tends to mark the first argument of both of these structures while accusative tends to mark the second argument of the two-place structure. This thesis will refer to such first arguments as subjects and second arguments as objects.

Following this, canonical argument structures are structures that incorporate subjects and objects. They are the predominant structure in many languages and

they entail structural cases. Non-canonical (oblique) argument structures entail semantic cases and tend to occur with fewer verbs than canonical structures.

### 2.1.6 Productivity of argument structures

In addition to canonical argument structures, a language may incorporate other argument structures that are used with fewer verbs (usually formed with semantic cases, rendering non-canonical argument structures). Hence, argument structures can be said to vary in terms of (syntactic) productivity (Barðdal, 2006, 2008; Zeldes, 2012). This means that they represent patterns, some of which are more ‘available’ to language speakers than others, therefore being assigned to a wider variety of verbs. Syntactic productivity is analogous to morphological productivity, where the availability of a derivational morpheme can be assessed in terms of various measures, e.g. the number of lemmas with which it can be found.

Similarly, Barðdal (2008) shows that in Icelandic, an argument structure with a nominative subject and an accusative object (3a) can be found with twice as many verb lemmas and is therefore twice as productive as a structure with a nominative subject and a dative object (3b):

- (3) a. Stormurinn blés strompinn af húsinu  
storm.NOM blew chimney.ACC off.house  
‘The storm blew the chimney off the house’
- b. Skipstjórinn sökk skipinu  
captain.NOM sank ship.DAT  
‘The captain sank the ship’  
(Svenonius, 2005)

Hence, highly productive structures are hereby referred to as canonical argument structures, which derives from the nature of morphological marking they use as well as their productivity. Less productive structures using non-canonical morphology may be referred to as oblique or non-canonical structures.

### 2.1.7 Argument status

In a binary view, the argument status of a constituent describes it either as an argument or an adjunct. In a gradient view, constituents may be said to be strong or weak arguments, or non-arguments (adjuncts).

Assumptions are often made about particular argument structures and the quality of argument status in them (Blake, 2001). Following the hypothesis that the main productive (canonical) argument structure must mark arguments (e.g. the non-marking structure in English, ‘I see you’), other properties of that structure have become extended as indicators of argument status in linguistic tradition, independently from the canonical structure to which they apply. For instance, the canonical structure in Icelandic marks objects in accusative, which is categorised

as a structural case. In addition, objects in this structure can also undergo the passive alternation, becoming nominative subjects (both features are widespread in Indo-European languages). Instead of regarding structural case selection and passivisation as properties of one particular (canonical) argument structure, they have come to be viewed as properties of arguments, describing the class in an exclusive manner. That is, the linguistic tradition has come to regard structural case and passivisation as prerequisites for being classified as an argument. The leap from regarding them as properties of one particular structure to viewing them as properties of arguments in general is often left without justification.

If the case and syntactic operations pertaining to one argument structure are viewed as prerequisites for constituting an argument, non-canonically marked arguments are automatically rendered qualitatively different from canonical arguments. For instance, both accusative case and passivisation apply to (3a), while neither applies to (3b). Dative complements retain their case in passives (Zaenen and Maling, 1984), marking the latter as an argument of a lesser type.

The present thesis does not make this assumption, nor does it assume a qualitative distinction between the quality of argument status marked by various structures. Instead, the general view about only one structure marking ‘true’ arguments is viewed as a plausible hypothesis that may be tested by experimental means (as is done in chapter 6). The rest of the studies presented in this thesis (chapters 4–5) are not about argument status.

## 2.2 Argumenthood indicators

The first of the three aims of this thesis is to “to investigate the distinctions between verbs in oblique and canonical argument structures with regard to lexical semantics and the usage patterns of verbs.” (chapter 1). For this, argument criteria are a useful starting point. Much of the literature mentions various definitive criteria which distinguish between arguments and adjuncts, including syntactic operations, semantic characteristics and morphological properties (Koenig et al., 2003; Arka, 2014; Forker, 2014; Ackema, 2015; Perek, 2015; Malchukov and Comrie, 2015; Barbu and Toivonen, 2016). Several of these criteria have been used to highlight the distinction between canonical and non-canonical argument structures, grouping the latter with adjuncts (Blake, 2001).

The set of proposed criteria is extremely varied. First, the set includes properties which pertain equally well to all argument structures, both canonical and oblique. Their described properties are so broad that they may be regarded as parts of the very definition of argument structures, contrasting one particular way a case is used to its other, non-structural uses. These include morphological selection, iteration, do-so replacement. They are outlined in section 2.2.1.

Second, the set of argument criteria includes features, which are highly problematic when viewed as properties of *all* arguments. They do not appear to account for the way argument structures (even all instances of the canonical struc-

ture) are expressed in language and many of them experience fundamental issues in a cross-linguistic perspective. They will be outlined in section 2.2.2.

Finally, for some features it has been shown that they distinguish between the ways various argument structures are used in a wide range of languages. They include semantic coherence and lexical transitivity. Section 2.2.3 will discuss these two criteria.

### 2.2.1 Criteria accommodating all argument structures

This section discusses three argument criteria which define an area so broad that all argument structures fit the description — morphological selection, iteration and do-so replacement. Given that they do not indicate meaningful distinctions between various structures, they may be more useful for outlining the notion of argument structure in general, rather than constitute variables that capture the variance between individual argument structures.

#### Morphological selection

The term ‘selection’ refers to the ability of a linguistic expression to impose restrictions on a co-occurring expression (Hole, 2015). Verbs select the morphological case of their arguments but not the case of adjuncts (Jacobs, 1994; Forker, 2014; Hole, 2015). Broadly, this criterion accommodates arguments in both canonical and non-canonical argument structures. However, the matter might be more complicated in individual languages.

Instead of only having verbs selecting one particular case for their arguments (e.g. the Estonian *sõltuma* ‘depend on’ always selects an elative complement), a language may have a differential case marking system. In such an instance, a verb may be said to select a (commonly narrow) range of cases. A vast amount of work suggests that in differential marking, case selection is affected by various semantic and pragmatic factors which have to do with subject and object properties as well as the lexical properties of the verb. Most may be described in terms of semantic/pragmatic transitivity (e.g. animacy, tense, aspect, mood) (Malchukov and De Swart, 2008; de Hoop and de Swart, 2009; Lestrade, 2010; Kittilä et al., 2011). For instance, Kittilä (2008) notes that a clause where the object of the Finnish verb *rakastaa* ‘love’ is in accusative case 1 is more transitive than when the object is in partitive case 2:

- a. Mies rakas-ti koira-n-sa kuoliaaksi  
man.NOM love-PST.3SG dog-ACC-3POS to.death  
‘The man loved his dog to death’
  
- b. Mies rakas-ti koira-a-nsa  
man.NOM love-PST.3SG dog-PAR-3POS  
‘The man loved his dog’  
(Kittilä, 2008, 360)

However, this does not pose an issue for the idea that morphological selection underlies all argument structures. The concept merely refers to both verbs selecting one particular case as well as verbs selecting a set of cases (e.g. *rakastaa* ‘love’ in 2.2.1).

## Iteration

Arguments cannot be iterated, meaning their number is limited in a clause. The number of adjuncts, however, is unlimited (Radford, 1988; Rákosi, 2006; Forker, 2014; Ackema, 2015). For instance, a clause can contain many types of temporal adjuncts (‘We will meet today at nine o’clock in the evening’), but the same is not true of arguments (\*‘I ate the chicken the wing’) (Rákosi, 2006, 102).

Essentially, what is meant by this constraint is that participant information is more restricted than the information expressed by adjuncts. This observation about participants underlies the whole concept of arguments. For instance, Fillmore (1977) posited that only a few participant roles in a given situation can be ‘put into perspective’ (Fillmore, 1977, 73). While ‘buy’ and ‘pay’ both activate the scene of a commercial event, the perspective of ‘buy’ includes the buyer and goods while the perspective of ‘pay’ includes the buyer and money. Similarly, Dowty noted that “We really only need two role types to describe argument selection efficiently” (Dowty, 1991, 571-572), referring to Proto-Agent and Proto-Patient (section 2.4 will further discuss Dowty’s Proto-roles in the context of argument status). Verbs with one or two arguments are pervasive while verbs with three arguments are rare and verbs with more arguments than three virtually non-existent (Ackema, 2015). All in all, an argument structure is commonly viewed as having a predefined number of argument slots, but it may combine with an unlimited number of adjuncts.

As to the question why argument number is limited, Lestrade (2010) regards the linguistic salience of two arguments as a result of event segmentation. People segment activities into events and subevents with hierarchical structure, in order to improve comprehension and optimise memory resources (Kurby and Zacks, 2008). In canonical transitive events, the event starts with an Agent instigating the event and ends with a Theme or Patient being affected by the event. In emotion and perception events or states-of-affairs, the event may start with a Stimulus and end with an effect on the Experiencer. Hence, two arguments are frequently salient as both segment and frame the event (Lestrade, 2010, 30).

In any case, similarly to morphological selection, this criterion classifies both canonical and non-canonical argument structures as structures marking arguments that are distinct from adjuncts.

## Do-so replacement

Ackema (2015, 260) proposes that the distinction between arguments and adjuncts is captured by do-so replacement, which must include the argument (4a)–(4b), but need not include the adjunct (4c):

- (4) a. John ate a banana and Geraldine did so, too (= ate a banana)  
b. \*John ate a banana, while Geraldine did so an apple (= ate x)  
c. John ate a banana yesterday, while Geraldine did so today

Do-so replacement broadly applies to Estonian arguments as well. The equivalent structure would be *tee-b seda* do-PRS.3SG this.PAR ‘does this’ as it is more common than *tee-b nii* do-PRS.3SG so ‘does so’.<sup>2</sup> Similarly to the English examples in (4), canonical objects in Estonian must be included in the replacement, as shown in (5):

- (5) a. Ma sõi-i-n omletti  
1SG.NOM eat-PST-1SG omlette.PAR  
‘I ate an omlette’  
b. Taavi teg-i seda ka  
Taavi.NOM do-PST.3SG this.PAR too  
‘Taavi did this, too’ (= ate an omlette)  
c. \*Taavi teg-i seda muna  
Taavi.NOM do-PST.3SG this.PAR egg.PAR  
‘Taavi did this an egg’ (= ate x)

Do-so replacement does not appear to be restricted to the canonical argument structure. It works well with Estonian arguments marked in spatial cases, which may not be omitted from the replacement (6):

- (6) a. Ma unista-n uue-st maja-st  
1SG.NOM dream-PRS.1SG new-ELA house-ELA  
‘I dream of a new house’  
b. Taavi tee-b seda ka  
Taavi.NOM do-PRS.3SG this.PAR too  
‘Taavi does this, too’ (= dreams about a new house)  
c. \*Taavi tee-b seda auto-st  
Taavi.NOM do-PRS.3SG this.PAR car-ELA  
‘Taavi does this a car’ (= dreams about x)

However, the replacement is restricted in terms of the lexical semantics of verbs, i.e. in terms of volitionality. The canonical argument structure may undergo do-so replacement if the subject is volitional (5), but not if the subject is non-volitional (7):

---

<sup>2</sup> However, the acceptability patterns of (5) apply with both structures.

- (7) a. Firma oma-b kah-te auto-t  
 company.NOM OWN-PRS.3SG TWO-PAR CAR-PAR  
 ‘The company owns two cars’  
 b. \*Keskool tee-b seda ka  
 high.school.NOM DO-PRS.3SG THIS-PAR TOO  
 ‘High school does this, too’ (= owns two cars)

The same applies to non-canonical argument structures (8):

- (8) a. Üritus sõltu-b ilma-st  
 event.NOM DEPEND-PRS.3SG WEATHER-ELA  
 ‘The event depends on the weather’  
 b. \*Piknik tee-b seda ka  
 picnic.NOM DO-PRS.3SG THIS-PAR TOO  
 ‘The picnic does this, too’

All in all, do-so replacement categorises some canonical and oblique arguments as arguments, but the replacement is semantically restricted. It does not work with verbs expressing state-of-affairs with non-volitional subjects, regardless of the argument structure occurring with the verb.

## 2.2.2 Highly problematic criteria

Opposite to the criteria in section 2.2.1 which define arguments in an all-inclusive manner, the criteria discussed in the present section are too problematic to be regarded as features pertaining to all arguments. It is most evident from the fact that they do not apply to all instances of the canonical structure. This seems to be due to the fact that they are either too simplistic with little cross-linguistic basis, or too restrictive, only applying to some instances of the canonical structure. They include syntactic obligatoriness, case, passivisation, morphosyntactic categories, semantic categories and semantic selection.

### Obligatoriness

The metaphor of arguments as slots in a formula contributes to an understanding that true arguments cannot be omitted from a clause. Advocating for a view where the notion of linguistic arguments are kept separate from event participants, Barbu and Toivonen (2016) point out that all eating verbs must entail food as a participant (a semantic argument), but not all of them encode food as a syntactic argument. The object in ‘Sally devoured the pizza’ must be realised for the clause to be acceptable, while the object in ‘Jasleen and Sam were dining’ has no such obligation. Hence, syntactically obligatory arguments are fundamentally distinct from non-obligatory arguments, to the point that only the former may be viewed as true arguments.

However, a qualitative distinction between syntactically obligatory and optional elements has proven difficult to establish as an argument criterion in a

cross-linguistic perspective. Languages vary widely in the ways arguments must or can be realised. For instance, languages encoding argument information on the verb via verbal agreement appear to be more prone to argument ellipsis. In Estonian, a morphologically rich language with verbs marked for the subject person and number, argument ellipsis is common for both canonically (Vihman, 2015; Erelt, 2017a) and non-canonically marked arguments (Lindström and Vihman, 2017; Lindström and Uibo, 2017). Hence, it appears that the notion of syntactic arguments works best for languages with impoverished morphology, where argument ellipsis is more rare, e.g. English.

Furthermore, it has been pointed out that argument ellipsis is at least partially semantically conditioned. Namely, the arguments of some predicates receive existential interpretation while others receive anaphoric interpretation (Fillmore, 1969; Fodor and Fodor, 1980; Condoravdi and Gawron, 1996):

- (9) a. There was a piece of bread on the table but John didn't eat.  
 b. There was a good job available here but Fred didn't apply.  
 (Condoravdi and Gawron, 1996, 3)

The implicit argument of 'eat' in (9a) is existential in the sense that according to this sentence, John did not eat anything (not just the bread). The implicit argument in (9b), however, is anaphoric, because it is the mentioned job to which Fred did not apply, but he may have well applied to other jobs. Verbs with existential implicit arguments (9a) allow them to be omitted (10a), while verbs with anaphoric implicit arguments do not (10b)<sup>3</sup> (Condoravdi and Gawron, 1996):

- (10) a. I painted last week.  
 b. \*I applied last week.  
 (Condoravdi and Gawron, 1996, 3)

Furthermore, not only does syntactic obligatoriness appear to depend on the semantics of verbs, it also appears to depend on the semantic qualities of objects, as shown in (11):

- (11) a. We can't afford to buy a house, so we have to rent.  
 b. \*He arrived without a car so he had to rent.  
 (Klotz, 2000, 13-14, via Hole 2015)

---

<sup>3</sup> In a way, this is linked to another argument criterion, i.e. latency, which states that when arguments are omitted from clauses, verbs can only be interpreted if the referent of the unexpressed argument is evident from discourse context (Croft, 2001; Forker, 2014). For instance, 'Joe won.' is only interpretable if the hearer knows what it is that Joe won from the context, meaning that the object of winning is a true argument. In contrast, temporal and spatial information can be easily identified by the hearer even if no reference to them is made in discourse context (Forker, 2014). This approach would still need to argue, why the object of 'paint' (10a) is less of an argument than the object of 'win' or 'apply' (10b).

Finally, viewing syntactic obligatoriness as a cross-linguistically valid criterion for arguments is especially controversial when it is evident that syntactic obligatoriness also affects a wide range of adjuncts. It is commonly assumed that adjuncts, e.g. information referring to the location, time, reason and manner of events (DeArmond and Hedberg, 1998), are peripheral to the predicate, meaning they cannot be obligatory in the clause in the way arguments are obligatory (Ackema, 2015). However, such phrases are obligatory in a number of syntactic contexts (Goldberg and Ackerman, 2001, 798), e.g. (12)–(15):

- (12) a. \*This book reads.  
 b. This book reads well.
- (13) a. \*Pat laughed a laugh.  
 b. Pat laughed a hearty laugh.
- (14) a. \*This house was built.  
 b. This house was built last year.
- (15) a. \*a built house  
 b. a recently built house

The contextually conditioned adjuncts in (12)–(15) have been explained semantically and pragmatically. Grimshaw and Vikner (1993) observe that adjuncts are obligatory for such accomplishment verbs (Vendler, 1957) as ‘build’ in (14), but not for other accomplishment verbs, such as ‘destroy’ (‘This house was destroyed.’). The distinction is explained semantically via event structure. Others have offered a pragmatic explanation for the examples (Goldberg and Ackerman, 2001; Ackema, 2015), suggesting that (12)–(15) are unacceptable without adjuncts because they lack information focus, meaning they fail to convey any new information without adjuncts (Goldberg and Ackerman, 2001).

In any case, most agree that the obligatory adjuncts in (12)–(15) do not constitute verbal arguments (Grimshaw and Vikner, 1993). Some, however, argue that they do:

- (16) a. \*Wir befind-en                    uns.  
           We be.situated-PRS.1PL REFL  
           \*‘We are situated.’
- b. Wir befind-en                    uns hier.  
           We be.situated-PRS.1PL REFL here  
           ‘We are situated here.’  
           (Müller, 2020, 34)

Müller (2020) analyses *hier* as an argument of the verb *sich befinden* in (16b), because the verb is unacceptable without the spatial phrase. Analogous examples include the arguments of ‘behave’ and ‘treat’ expressing manner (‘behave well’, ‘treat badly’) (Forker, 2014, 29), as well as the spatial Goal of ‘put’ (\*‘Beatrice put her bike.’). Similarly to more canonical arguments, such phrases are semantically,

morphologically and syntactically restricted when occurring with these particular verbs. In addition, the time and location phrases in ‘lasted three hours’ and ‘went to New York’ have also been analysed as semiarguments (Matthews, 1981, 141). Section 2.4.2 will further discuss approaches with a graded view on argument status.

Finally, there is some evidence that participant locations (locations only describing one of the event participants, e.g. ‘I hid the wallet **under the pillow**’) and Instruments (‘I hit it **with a hammer**’) constitute semi-arguments to some verbs. Koenig et al. (2003) showed that speakers associate verbs with participant locations (‘I boiled eggs **in a pot**’) more strongly than with event locations describing all participants (‘I boiled eggs **in the kitchen**’). Similarly, Instruments are more strongly linked to some verbs (‘slice’) than others (‘wash’) (Koenig et al., 2003; Rissman et al., 2015).

In conclusion, viewing constituents that may be omitted as weaker arguments than those which must overtly be expressed (Barbu and Toivonen, 2016) and viewing syntactic obligatoriness as a property of prototypical argumenthood (Somers, 1984; Ágel, 2000; Arka, 2014; Forker, 2014) stand on rather shaky basis. First, ellipsis is affected by the morphosyntax of a language as well as the semantics of verbs and complements. Second, adjuncts can be obligatory as well. However, such a criterion may well pertain to a single language. In order to show this, one ought to provide an overview of the semantic and pragmatic conditions facilitating syntactic obligatoriness and show that the phenomenon indeed applies widely enough to align with morphosyntactic devices marking argument structure.

Cross-linguistically we may assume that syntactic obligatoriness is a phenomenon parallel to argumenthood and adjuncthood rather than complementary to them. It appears to affect both types of elements idiosyncratically in each language. Hence, the current thesis will regard both covert and overt arguments are analogous in terms of argumenthood status (Tesnière, 2015; Ackema, 2015; Hole, 2015).

## Case

Arguments are expected to be marked in structural case and adjuncts in semantic case (Blake, 2001). The idea that some cases universally classify as structural (syntactic, direct, grammatical), marking verbal arguments, while others are semantic (oblique<sup>4</sup>, local), marking non-arguments, underlies the Indo-European linguistic tradition. It was already reflected in the works of Jakobson (1936), who was concerned with establishing the ‘overall meaning’ (*Gesamtbedeutung*) of each case. In the recent decades, finding the single meaning of affixes has not been as important in linguistic research. Regardless, the assumption persists

---

<sup>4</sup> In some approaches, ‘oblique’ in case description refers to the distinction between nominative (*orthē*, ‘straight’) and all other cases (*ptōseis plagiai* ‘oblique’ or ‘slanting’) (Blake, 2001)

(overall in modern linguistics as well as in Estonian linguistics), that cases may be divided into separate, discrete categories (Blake, 2001; Viht and Habicht, 2019).

In accusative-nominative languages, the category of structural cases is often viewed as including nominative, accusative and/or genitive (Jakobson, 1936; Kuryłowicz, 1964; Blake, 2001). Cases in this view are inherently linked to syntactic relations, subjects being linked to nominative and objects to accusative. However, the particular contents of the structural and semantic case categories are generally believed to depend on the language. For instance, structural cases in Estonian include the three cases participating in the differential argument marking system — nominative, partitive and genitive (Erelt and Metslang, 2017; Viht and Habicht, 2019).

Given that the system relies on cases being divided into groups based on their functions, its applicability depends on the degree of multifunctionality in the case system. Although semantic cases ought to primarily mark adjuncts, they have been noted to occur in argument structures (both cross-linguistically as well as in Estonian). Several authors have argued that complements marked in what are traditionally seen as semantic cases still qualify as arguments of the verb. Kuryłowicz (1964) notes that in a governed position, semantic cases (in his approach, dative, ablative, instrumental and locative) are to be considered “case forms and not as adverbs, just because under certain conditions their endings are apt to become mere signs of syntactical dependence, i.e. allomorphs of the acc. of direct object.” (Kuryłowicz, 1964, 193).

Nichols (1983; 1984) takes this idea further, saying that *because* arguments in semantic cases qualify as fully functional arguments, analogous to those marked in structural cases, one ought not to assume a universal functional distinction between structural and semantic cases. Comparing Nanai, Old Russian, Russian, Chechen-Ingush and Finnish, she shows that structural cases are functionally distinct from semantic cases in some languages, while in others they are not. “Case functions differ widely, but they turn out to fall into two basic types, which I call overlapping and complementary.” (Nichols, 1983, 181). Not only do Russian and Finnish have cases which mark both arguments and non-arguments, but this applies to most of their cases. In Chechen-Ingush and Nanai, however, only very marginal functional overlap is found in the case system. Hence, some languages have structural and semantic case categories, while others do not. In addition, Nichols (1983) showed that a language with grammaticalised core arguments (meaning the language has a predominant argument structure, involved with particular morphosyntactic processes) may still have functionally overlapping cases. For instance, various syntactic rules apply only to subjects and objects in Russian, but five of its six cases are functionally ‘overlapping’.

Hence, Nichols (1984) criticised the case-based approach to argumenthood, positing that one ought not to use case to distinguish between arguments and adjuncts. She suggested that the reason why such a practice, i.e. omitting oblique

arguments in two-place verbs from the argument group, is nevertheless widespread in linguistic tradition, is because of historical accident, brought about by several factors. First, the habit of traditional grammatical descriptions of characterising theoretical notions with the aid of examples has led modern linguistics to be overly concerned with prototypical instances of notions (productive argument structures), at the expense of their peripheral variants (less productive structures). Second, the morphosyntactic patterns particular to English, where obliques surface as prepositional phrases (‘depends **on the weather**’) has also contributed to a bifurcated system of cases (Nichols, 1984, 184).

All in all, similarly to syntactic obligatoriness, it is difficult to maintain the cross-linguistic validity of the idea that case is informative of argument status and that only some cross-linguistically defined cases have the function of marking arguments. Nichols (1983) suggests that case function overlap is a typological feature, which pertains to some languages and not to others. As will be shown in section 2.3, the grammaticalisation perspective of cases further complicates this matter.

### Passivisation

In addition to being marked by a structural case, the second arguments of two-place verbs are defined by their ability to become nominative subjects in passive alternation (Blake, 2001; Ackema, 2015). This property does not often extend to oblique arguments (Blake, 2001, 34), which either retain their case in passive sentences (as with dative complements in Icelandic, see Zaenen and Maling, 1984; Svenonius, 2005) or merely cannot occur in passives:

- (17) a. They built the houses in 1959  
 b. The houses were built in 1959  
 c. \*In 1959 was built the houses  
 d. Our trip depends on the weather  
 e. \*The weather was depended (on) (by our trip)  
 (a-c originate from Ackema, 2015, 268)

There are several issues related to passivisation. First, similarly to do-so replacement, the process is semantically restricted, meaning it only applies to a subset of verbs using the canonical argument structure. For instance, stative predicates with non-volitional subjects rarely passivise (Haspelmath, 2003; Haiden, 2005), as is shown with the Estonian verb *omama* ‘own’, selecting a canonical object<sup>5</sup>:

- (18) a. Ma oma-n kah-te auto-t  
 1SG.NOM OWN-PRS.1SG TWO-PAR CAR-PAR  
 ‘I own two cars’

<sup>5</sup> Some Estonian emotion verbs nevertheless occur in passives, e.g. *ole-n kõigi poolt karde-tud* COP-PRS.1SG all.GEN POST fear-PPP ‘I am feared by all’ (Erelt, 2017b, 220)

- b. \*Kaks auto-t on oma-tud (minu poolt)  
 two.NOM car-PAR AUX OWN-PPP 1SG.GEN POST  
 ‘Two cars are owned (by me)’
- c. \*Kaks auto-t on (minu) oma-tud  
 two.NOM car-PAR AUX 1SG.GEN OWN-PPP  
 ‘Two cars are owned (by me)’

Furthermore, this restriction also applies when stative interpretation is of pragmatic origin. For instance, a number of verbs alternate between causative active and non-causative stative interpretations, depending on the semantics of their complements (Rothmayr, 2009). The Estonian *piirama* ‘surround’ is active in (19) and stative in (20). Passivisation is only available to its active interpretation:

- (19) a. Sõjavägi piira-b linna  
 army.NOM besiege-PRS.3SG city.PAR  
 ‘Army besieges the city’
- b. Linn on sõjaväe poolt piira-tud  
 city.NOM AUX.PRS.3SG army.GEN POST besiege-PPP  
 ‘City is besieged by the army’/‘The city is under siege by the army’
- (20) a. Mäe-d piira-vad järve  
 mountain-NOM.PL surround-PRS.3PL lake.PAR  
 ‘Mountains surround the lake’
- b. ??Järv on mägede poolt piira-tud  
 Lake.NOM AUX.PRS.3SG mountain-GEN.PL POST surround-PPP  
 ‘The lake is surrounded by the mountains’

In addition to semantics or pragmatics, passivisation also depends on argument structure. The Estonian verb *virutama* ‘hit’ takes an oblique argument in allative case (‘onto’). Even though it expresses a highly transitive, dynamic event, its passive form is unacceptable (21b)–(21c), even when the oblique argument is allowed to retain its case in subject position (21b):

- (21) a. Laps viruta-s koera-le  
 child.NOM hit-PST.3SG dog-ALL  
 ‘The child hit the dog’
- b. \*Koera-le on viruta-tud lapse poolt  
 dog-ALL AUX.PRS.3SG hit-PPP child.GEN POST  
 ‘The dog was hit by the child’
- c. \*Koer on viruta-tud lapse poolt  
 dog.NOM AUX.PRS.3SG hit-PPP child.GEN POST  
 ‘The dog was hit by the child’

Passivisation is also highly language-specific. For instance, Icelandic allows oblique arguments to be fronted in passives if they retain their case (Zaenen and Maling, 1984), but this is unacceptable in Estonian (21b). In addition, restrictions on passive may well be affected by other valency-changing operations available in a language. For instance, differently from Icelandic, Estonian has a productive

system of impersonal voice with a fronted object retaining its original case and an elided subject, which might lead to a more restricted passive.

All in all, passivisation is a complex phenomenon with wide cross-linguistic variation. It can be restricted by the stativity of events and states described by verbs and the non-volitionality of their subjects, as well as argument structures. Hence, it only appears to describe one argument structure, and only a subset of verbs occurring with it. If no distinction is assumed to exist between the argument status marked by canonical and non-canonical argument structures, passivisation merely remains an idiosyncratic operation pertaining to one argument structure, unconnected to the notion of argument status. This distinction is in the focus of the experiment laid out in chapter 6.

### Morphosyntactic categories

This criterion proposes that arguments surface as noun phrases while adjuncts tend to be adpositional phrases (Koenig et al., 2003; Perek, 2015; Malchukov and Comrie, 2015). The notion has received much criticism (Nichols, 1984; Hoekstra, 2004; Ackema, 2015). First, languages vary widely in their use of morphological cases and adpositions. In a sample of 261 languages, 100 languages were found to have no morphological case marking (Iggesen, 2013). It is therefore difficult to exclusively link case (rather than adpositions) to such highly predominant roles as arguments.

Second, Ackema (2015, 266) notes that both arguments and adjuncts can be realised as embedded clauses in English (**‘That smoking is bad for your health worries most people’**). In addition, adjuncts can be noun phrases (**‘Harry met Sally the other day’**). Furthermore, Ackema notes that oblique arguments are expressed with prepositions in English (**‘I am counting on you’**), but should nevertheless be interpreted as arguments. This is because the selection of the preposition ‘on’ by the verb is not semantically motivated (there is no spatial component to the concept of counting on someone). In addition, its do-so replacement version is relatively acceptable (Ackema, 2015, 267), as was also shown with Estonian obliques in (4c). All in all, similarly to obligatoriness and case, this criterion does not pattern well with the notion of argumenthood.

### Semantic categories

Certain semantic roles are linked to arguments while others are regarded as adjunct-related. Arguments have been linked to Theme, Patient, Goal, Source, Path, Experiencer roles, while adjuncts are linked to Time, Location, Manner and Reason (DeArmond and Hedberg, 1998).

Some approaches use role hierarchies to link argument positions to particular roles in a direct manner. Hierarchy-based role systems have been extensively criticised (Levin and Rappaport-Hovav, 2005; Newmeyer, Frederick, 2010; Kittilä

et al., 2011) as there is an abundance of research showing that mapping information to grammatical relations involves a wide variety of factors uncaptured by such linear systems (Levin and Rappaport-Hovav, 2005).

Furthermore, experimental approaches have shown that while broad assumptions about semantic categories appear to hold (Koenig et al., 2003), argumenthood may be inherently gradient. In such a view, some categories fall in between quintessential arguments and adjuncts. For instance, participant locations and Instruments have been shown constitute stronger arguments than time and manner phrases (Koenig et al., 2003; Rissman et al., 2015). Koenig et al. (2003) concluded that “[...] argument status cannot be adjudicated on the basis of participant category. Constituents expressing instrument participants are neither uniformly adjuncts nor arguments.” (Koenig et al., 2003, 96). In any case, as a criterion, the notion of semantic categories appears to be problematic.

### **Semantic selection**

Verbs are said to impose semantic restrictions on arguments (Hole, 2015). Semantic selection or the co-occurrence criterion (Forker, 2014) means that ‘eat’ selects tangible objects, while ‘think through’ selects more abstract objects (Hole, 2015). Adjuncts, on the other hand, are not selected by verbs, resulting in more variation in their semantics. Instead, adjuncts are said to select (modify) verb phrases, e.g. ‘for Lisa’ only attaches to volitional events (Hole, 2015).

There are several issues related to semantic selection. First, the direction of selection is not clear and one could also posit that arguments select the events that could happen to them. For instance, ‘our family dog’ can select ‘pat’ as an event by which to be affected, but not ‘program’, because only inanimate objects can be programmed (Hole, 2015).

Second, verbs can be highly polysemous, but even relatively monosemous verbs occur with a wide variety of objects. Abstract verbs can accommodate virtually all types of complements (e.g. ‘My trip depends on my mum/the weather/good fortune/my dog’). Many concrete verbs may do the same by rendering metaphorical interpretations (‘The tornado destroyed my house/my mum/my business’).

Finally, it is not clear that selection pertains only to arguments, because verbs also impose semantic restrictions on various other types of information. For instance, the event of eating may only occur in certain types of manner and not others (‘I ate my sandwich hard.’). In summary, semantic selection appears to describe a phenomenon outside the argument-adjunct distinction.

#### **2.2.3 Criteria describing the variation between argument structures**

Finally, there are two criteria which appear to capture a significant degree of the potential variance between various argument structures — semantic coherence and the lexical transitivity of verbs. Both Blake (2001) and Nichols (1983) note

that these two properties pertain to canonical structures and not to oblique structures, even though they disagree in what type of argument status is coded by these structures. Both criteria describe different aspects of the meaning of the argument structure expression and they may explain why some verbs are more likely than other verbs to be assigned non-canonical structures.

### Semantic coherence

Some verbs describe events where participants are involved in highly generalised ways, with no spatial, temporal or otherwise semantically coherent (or specific) relationship. For instance, the partitive case on the object of the Estonian verb *armastama* ‘love’ only conveys that the entity described by the object participates in the state-of-affairs of someone else loving it. The complement of *lähenema* ‘approach’, however, is marked in allative (‘onto’), the selection of which is likely motivated by allative being a spatial case marking Goals. The entity being approached by something else would qualify as a Goal.

Crucially, the structural use of semantic case is not thought to be as non-coherent as that of structural case (Nichols, 1983; Blake, 2001). When a verb selects a semantic case for its argument, the selection is thought to be motivated by the original semantics of the case. Blake (2001, 34) notes that ablative is governed by the Latin verb *ūtī* ‘use’, but Latin ablative also expresses instrumental function in other clauses. Hence, its government by the verb *ūtī* is semantically motivated, because the meaning of ‘use’ is highly semantically compatible with instrumental function (use a sword).

Blake (2001) presents further evidence. He argues that when a verb occurs with both a canonical and a non-canonical argument structure, these two structures select different types complements, further demonstrating the fundamental distinction in the argument status marked by canonical and non-canonical structures. The Latin verb *moderārī* ‘govern’ selects dative when referring to self-restraint (‘govern **one’s anger**’) and accusative when referring to external elements (‘govern **territory**’).

Nichols (1983) agrees that the structural use of semantic case may include some of the original semantics of that case (calling this ‘direct semanticity’). The selected case may directly index the semantic role of the complement or the lexical semantics of the verb, which is why she calls the effect ‘direct semanticity’ (Nichols, 1983, 186). However, she does not agree that coherent use of semantic case automatically indicates weaker or lower argument status. She posits that a constituent may still constitute a core argument of that verb, even if the relation between the two is semantically linked to the original meaning of the semantic case.

All in all, both approaches assume that ‘direct semanticity’ is present in oblique structures and not in canonical structures. Because this underlies the conviction of one of them, that this excludes oblique arguments from the set of true

arguments, it is an observation with potentially profound effect. However, it remains a hypothesis, because no empirical evidence exists to demonstrate such a tendency in oblique argument verbs. Essentially this is a quantitative question, asking: what proportion of verbs occurring with oblique structures encode traces of the original meaning of the case?

There is reason to believe that the effect is not as all-encompassing among oblique argument verbs as is suggested. It has been noted that semantic cases may be used without bearing any semantic traces of their original meaning (Kuryłowicz, 1964; Nichols, 1983, 1984; Ackema, 2015). For instance, the Estonian elative case ('out of') has been noted to be highly multifunctional (Nurka, 2014; Viht and Habicht, 2019). In some contexts it appears to index some degree of its original semantics (22a). In others, however, it appears to mark event participation without any spatial meaning (22b):

- (22) a. Ma järelda-n su ilme-st, et ole-d  
 1SG.NOM conclude-PRS.1SG 2SG.GEN expression-ELA that COP-PRS.2SG  
 väsinud  
 tired  
 'I conclude from your expression that you are tired'
- b. Ma hooli-n sinu-st  
 1SG.NOM care-PRS.1SG 2SG-ELA  
 'I care about you'

All in all, the hypothesis that some argument structures persistently entail a higher degree of coherent semantics than other structures may be tested in a quantitative sample of verbs occurring with different types of structures. Such an investigation is undertaken in chapter 4. This assessment would contribute to both the understanding of the argument structure system in a single language, as well as be cross-linguistically relevant, describing the phenomenon of oblique argument structures in general.

### Lexical transitivity

In addition to semantic coherence, Nichols (1983) and Blake (2001) agree that verbs occurring with the canonical argument structure tend to be more semantically or pragmatically transitive than verbs occurring with non-canonical argument structures. Blake (2001, 34) notes that "Verbs taking peripheral complements are not activity verbs involving impingement on a patient, though this is not to imply that all verbs with accusative-marked direct objects are such verbs." However, the two authors again differ in what they think this means for argument status. While Blake views the lack of transitivity in oblique verbs as an indicator of a lesser type of argument status, Nichols (1983), views this tendency ('indirect semanticity' (Nichols, 1983, 186)) as not meaningful for argument status. In her view, less transitive verbs may nevertheless select arguments analogous to more transitive verbs (Nichols, 1983).

The effect of transitivity on argument structure has received relatively more attention than the presence of the original semantics of case (semantic coherence). As a semantic-pragmatic property of clauses, transitivity has been shown to extensively affect differential case marking (Onishi, 2001). In addition, it has been shown that when verbs select one particular oblique case with no alternation ('split alternations' in (Malchukov, 2005)), the lexical transitivity of verbs plays an important role in determining which verbs would do that.

Hence, verbs selecting oblique arguments are generally associated with decreased transitivity (Nichols, 1984; Onishi, 2001; de Hoop and Malchukov, 2007; Malchukov and De Swart, 2008; Kittilä, 2008; Lestrade, 2010). Decreased transitivity implies increased stativity. This means that a verb may come to be used with a non-canonical argument structure if it expresses a more stative state-of-affairs, e.g. when it encodes an Experiencer (e.g. 'be sad') or Possessor ('have') (Malchukov, 2005; Rákosi, 2006; Rice and Kabata, 2007; Erelt and Metslang, 2008; Lindström and Vihman, 2017).

This also means that oblique argument structures may become linked to particular verb types with decreased transitivity, such as modal verbs ('need') and verbs expressing occurrences ('drift') (Onishi, 2001, 25). Tsunoda (1981) proposed a more detailed verb type hierarchy to organise the observed patterns of co-occurrence. In an elaborated version of this hierarchy, Malchukov (2005) postulates that some verb classes select oblique arguments due to the decreased subjecthood of Agents (e.g. perception and emotion verbs), while others select them because of the decreased patienthood of Themes or Patients (e.g. pursuit verbs) (Malchukov and De Swart, 2008, 343). These verb hierarchies are broadly based on various semantic features indicating transitivity (Hopper and Thompson, 1980; Tsunoda, 1981), as well as the Proto-Agent and Proto-Patient properties in (Dowty, 1991) (section 2.4.1).

All in all, if we aim to map the characteristics of oblique argument structures which might distinguish them from canonical argument structures, lexical transitivity of verbs is a useful starting point. The hypothesis that argument structures differ in terms of verb transitivity is widely discussed (Onishi, 2001; Malchukov, 2005), but it lacks quantitative evidence from a wider range of languages. Investigating it by means of comparing a sample of verbs in oblique structures to those in canonical structures would be highly informative for understanding the dynamics which facilitate the assignment of argument structures in a language. In addition, quantitative evidence on the transitivity distinctions between verbs in two types of structures would be relevant to the broader cross-linguistic discussion on oblique argument structures.

## 2.2.4 Summary

This section discussed eleven variables that have been used as definitive descriptive features of arguments. Three were shown to broadly describe the concept of

argument structure (morphological selection, iteration, do-so replacement), pertaining equally well (or equally badly) to all argument structures. However, it is difficult to operationalise them for a comparative perspective, because they appear to show no relevant distinctions between individual structures.

Six of them were shown to experience a range of issues, many describing phenomena which do not pattern with the range of observed linguistic data. They include obligatoriness, case, passivisation, morphosyntactic categories, semantic categories and semantic selection. They were deemed to not constitute a useful basis for describing the system of argument structures in a language.

Finally, two semantic aspects of argument structures were highlighted as describing features in terms of which various argument structures may be expected to vary — the degree of semantic cohesion across distinct case functions or senses, and the degree of lexical transitivity in a verb. Both make assumptions about the division of labour inside a system of argument structures in language, as well as about the factors facilitating the assignment of structures to verbs.

Semantic cohesion also links this line of investigation to another framework, i.e. grammaticalisation. As is discussed in section 2.3, the potential of some semantic cases to occur with less coherent meaning than others appears to be linked to their degree of grammaticalisation. This means that both the synchronic and the diachronic perspective individually only tell half the story. The present theory argues that in order to provide a comprehensive description of the way a case system is used in the argument structure system of a language, one must join the two perspectives together. The next section will discuss the grammaticalisation of cases and why it is relevant to argument structures.

### **2.3 Cases as a dynamic system**

Semantic cases have been noted to occur in contexts highly similar to those of structural cases. They may be selected by verbs and they may occur with virtually no trace of their original semantics, merely marking a relationship of event participation. This functionality may be expected to be an expression of a higher degree of grammaticalisation for a semantic case.

Grammaticalisation is a process by which a linguistic element (lexeme, affix, etc) gains (more) grammatical functions (Lehmann, 1985; Heine and Kuteva, 2002). It is assumed to apply in a unidirectional manner, i.e. it is believed to be rare to observe a functional element becoming less grammatical over time. Cases themselves are the result of grammaticalisation. Using morphological cases means encoding meaning by means of a limited set of semantically condensed suffixes, which have evolved from other structures over time.

With cases, grammaticalisation has been observed to follow a particular type of cline (Bybee, 1985; Lehmann, 1985; Hopper and Traugott, 2003; Heine, 2008; Lestrade, 2010):

(23) Noun/verb → Spatial adposition → Locative case → Structural case  
(Lestrade, 2010, 2)

Each category on the cline has been noted to develop from the category to its left, the degree of grammaticalisation increasing from left to right. This is also reflected in the fact that type count decreases from left to right, while token count increases (Lestrade, 2010). This means that the class of nouns includes many more types than the class of adpositions, while the number of different structural cases is the smallest. Token count distinction is also reflected per each individual case, because each structural case may be assumed to have more tokens than each semantic case, individual adposition, noun and verb, in corpus data.

In a very broad perspective, two processes accompany the grammaticalisation of cases — their meaning becomes more generalised or ‘bleached’ and their form becomes shorter. Hence, as semantic cases become more structural, their original meanings become more abstract and semantically less coherent (Barðdal, 2008) (which is sometimes referred to as ‘semantic extension’ or ‘desemanticisation’). Hence, the more grammaticalised the case, the more bleached functionality may be expected of it. Participating in an argument structure may also be viewed as an indicator of some degree of grammaticalisation on the part of the case, because this function is often conceptualised as highly semantically bleached.

The cline merely sketches out a tendency, rather than presents a system with predictive power. That is, there are examples of grammatical change that do not follow the flowchart in (23). For instance, adpositions may gain a structural function without becoming semantic case affixes first, as was the case with the Latin spatial preposition *ad* (‘at’/‘towards’), which grammaticalised into a marker of indirect objects (Lehmann, 1985, 7). In addition, markers may only become partially functional in any particular step. For instance, semantic case affixes may give rise to structural cases, the use of which is conditioned by various semantic properties of these nominals. Multifunctional ‘semantic’ cases which have not evolved further towards the structural realm in a long while and merely maintain their wide range of functions also complicate the simplistic sketch in (23) (Nichols and Timberlake, 1991).

### 2.3.1 The origin of semantic cases

Nevertheless, many instances have been observed where semantic cases constitute new forms of older adpositions with somewhat similar phonology and meaning. The adpositions which give rise to semantic cases, in their turn, originate from nouns and verbs, the latter being a more frequent source Blake (2001). Nouns vulnerable to this change have been noted to include relator nouns (‘top’, ‘edge’),

body parts ('head', 'ear'), environmental landmarks ('sky') and abstract spatial notions ('length') (Svorou, 1994; Blake, 2001). For verbs, however, predicates such as 'come' and 'leave' may become spatial adpositions when one-place predicates ('fly') need to be used with a Goal or a Source, resulting in serial verb constructions ('fly leave' expressing 'fly to' in Thai). Similarly, verbs like 'give' become adpositions marking Recipients and verbs like 'take' come to mark instruments (Blake, 2001, 63).

However, adpositions becoming case affixes implies more profound change than nouns or verbs becoming adpositions. While both steps likely entail phonological change with lexemes/adpositions becoming shorter and less prominent than their sources (Harris and Campbell, 1995), becoming an affix normally involves becoming unstressed, vowels often becoming centred to a schwa (Blake, 2001). In addition, they are likely to develop several phonological variants (e.g. with vowel harmony), depending on the phonology of the host lexeme. This, in turn, facilitates the emergence of inflectional classes. For instance, ergative in Djaru (Pama–Nyungan, Australia) is realised as *-lu* when affixed on vowel-final stems and *-tu* on consonant-final stems (Blake, 2001, 166).

Postpositions are especially prone to becoming semantic case affixes (Harris and Campbell, 1995; Kulikov, 2008). For instance, the Estonian comitative suffix *-ga* originates from the comitative postposition *kaas* used in the 16th and 17th centuries (Rätsep, 1979; Aigro, 2020). However, in rare cases, prepositions may also give rise to case prefixes (Kulikov, 2008, 444).

When adpositions become case affixes, they undergo semantic change. Generally this means their new form will cover a broader semantic range than their old form as it becomes bleached and less coherent. First, this is evident from the difference between case systems and adposition systems. Namely, case affixes are less numerous than adpositions, rarely exceeding 10. In a sample of 261 language, Iggesen (2013) found only 24 languages (9.2%) with more than 10 cases. Adpositions, on the other hand, is a class that often includes a wider range of elements. This means that each case in itself must be more bleached and multi-functional than the adpositions from which they evolved, because it will have to cover a wider range of distinct meanings.

This distinction in functionality is evident in languages where spatial postpositions are used in parallel with nearly synonymous spatial cases. For instance, Estonian has both an adessive case ('on') (24a) and an adessive postposition ('peal') (24b):

- (24) a. Telefon on laua-l  
 phone.NOM COP.PRS.3SG table-ADE  
 'The phone is on the table'
- b. Telefon on laua peal  
 phone.NOM COP.PRS.3SG table.GEN POST  
 'The phone is on the table'

Klavan et al. (2011) has shown that in such contexts the semantic space covered by the affix is wider than the space covered by the adposition. This is evident from the fact that when the relation with the spatial element is more abstract and space semantics is metaphorical (e.g. *mure on südame-l* worry COP.PRS.3SG heart-ADE ‘Worry is on the heart’/‘I worry’), speakers prefer to use adessive case rather than the postposition *peal* ‘on’ (Klavan et al., 2011; Klavan, 2012).

### 2.3.2 From semantic cases to structural cases

When a semantic case becomes a structural case, further morphophonemic reduction (erosion) is expected, as well as further semantic bleaching. A case is functionally and semantically extended, “giving rise to case polysemy” (Heine, 2008, 466).

As mentioned, the cline in (23) is merely a helpful sketch, very broadly illustrating that a pathway exists between semantic and structural cases. In reality, the process is anything but simple. Nichols and Timberlake (1991) investigate the structural extension of the instrumental case in Russian. They demonstrate that neither of the presuppositions of the simplistic scenario of grammaticalisation hold — not the fact that grammatical phenomena are clearly distinct from non-grammatical phenomena, nor the fact that change always leads to entropy. Instead, they observe that the Russian instrumental has been a multifunctional case for a long while and many of the changes in its functions may be described as lateral, rather than teleologically leading to a state of higher grammaticalisation. They conclude that the case is simultaneously grammaticalised and fluid in every separate diachronic stage they observed (Nichols and Timberlake, 1991, 143). This means to illustrate that instead of a relatively quick change between functionally simple states, one may observe highly multifunctional cases instead, which resist simple categorisation and show no intention of decreasing their number of functions.

As this thesis focuses on spatial cases, the following sections will describe the pathway between spatial cases and structural cases from the perspective of two different directions — lative (‘onto’/‘into’) and separative (‘from on top of’/‘out of’). Locative configuration has been discussed less as a structural case source than the two directional configurations.

#### Lative affixes as structural case sources

The grammaticalisation of semantic case into structural case is often illustrated with the link between allative (‘onto’) and Experiencer/Possessor and Recipient arguments (Heine and Kuteva, 2002; Rice and Kabata, 2007; Lindström and Vihman, 2017). Goal markers (allatives) expressing motion to or towards a place have been noted to be cross-linguistically highly polysemous. Via metaphorical extension, they tend to develop DATIVE, BENEFACTIVE and PURPOSIVE functions as

an abstraction of their spatial sense (Rice and Kabata, 2007). In their study of 54 allative markers from 44 languages, Rice and Kabata (2007) map 33 different meanings for these markers, with a third of the morphemes having RECIPIENT function and another third CONCEPTUAL (Experiencer) function.

Several hierarchical systems have been proposed for the description of the semantic bleaching of allative case. Based on a study involving 77 languages, Blansitt (1988) proposes the following hierarchy:

- (25) Object ← Dative (indirect object) ← Allative ← Locative  
(Blansitt, 1988, 177)

The hierarchy in (25) is viewed as having predictive nature. It suggests that if one case affix marks both objects and allative elements, it also marks indirect objects (dative). If the same affix marks indirect objects and locative (non-directional spatial) meaning, it also marks allative (directional) meaning.

Similarly, Heine (1990) noted that the semantic extensions involving allative markers are not coincidental, but follow a systematic pathway, leading further away from concrete spatial meaning. For instance, its stative spatial function (LOCATION) gives rise to TIME and POSSESSION functions (Heine, 1990, 131). Two of these functions are represented in the Estonian allative marker *-le* which has the LOCATION (26a) as well as the POSSESSION function (26b):

- (26) a. Merili toetu-s aia-le  
Merili.NOM lean-PST.3SG fence-ALL  
'Merili leaned on the fence'  
b. Jakk kuulu-b Merili-le  
jacket.NOM belong-PRS.3SG Merili-ALL  
'The jacket belongs to Merili'

According to Heine (1990), the more dynamic lative meaning of allative is extended in a similar fashion, leading to PURPOSE, REASON and MANNER functions on the one hand, and RECIPIENT and BENEFICIARY functions on the other hand. The latter two also pertain to the Estonian allative suffix, as in (27a)–(27b):

- (27) a. Merili and-is Mari-le kommi  
Merili.NOM give-PST.3SG Mari-ALL candy-PAR  
'Merili gave Mari candy'  
b. Merili teg-i Mari-le heateo  
Merili.NOM do-PST.3SG Mari-ALL favour.GEN  
'Merili did Mari a favour'

However, the hierarchy in (Heine et al., 1991) is missing the EXPERIENCER function as well as entirely bleached and object-like THEME function, which are also seen with the Estonian allative in (28a)–(28b).

- (28) a. Merili-le meeldi-b Mari  
 Merili-ALL like-PRS.3SG Mari.NOM  
 ‘Merili likes Mari’
- b. Teispäev järgne-b esmaspäeva-le  
 Tuesday.NOM follow-PRS.3SG Monday-ALL  
 ‘Tuesday follows Monday’

All in all, spatial meaning can become entirely bleached in the lative dimension. As a case suffix obtains new functions, the old functions may remain in place, resulting in a highly multifunctional affix. Due to this multifunctionality, it is difficult to view such affixes as mere starting points or end points in a grammaticalisation process (Nichols and Timberlake, 1991).

### Separative affixes as structural case sources

While allative and dative are the two most widely discussed cases from which syntactic markers have arisen (and the only mentioned source categories in Heine and Kuteva, 2002), other parts of the spatial domain, too, have been noted to be productive sources for syntactic markers. For instance, separative case, i.e. the semantic opposite of lative case, is the source category of the Estonian partitive case, the dominant object case participating in the Estonian differential object marking system (Rätsep, 1979). Heine (2008) notes that ablatives (with separative meaning) are among the most common sources for partitives.

In Estonian, it is likely that the separative meaning of the source affix became bleached in the following way. All cases with a specific semantically coherent meaning can be expected to have the potential to accommodate abstract nouns, resulting in metaphorical interpretations of case meaning (‘in my mind’). Separative affixes may therefore express both spatial meaning (‘from the house’) as well as metaphorically spatial meaning (‘from this set/group’). The latter, however, is essentially what is meant by partitive meaning (‘two of these houses’ ~ ‘two from among these houses’ where the set of houses constitutes an entity from which something emerges). With the source of Estonian partitive, both functions were likely marked by the same case affix.

Partitive meaning, referring to a relationship where something is part of something else, became more bleached, but its atelic aspect was retained and it came to be used to refer to unfinished events (Larjavaara, 1991). This is the role partitive most prominently plays in the Modern Estonian differential object marking paradigm. It mostly marks objects which are affected to an unspecified degree. Miljan (2009) shows that verbs exclusively selecting partitive objects, e.g. *uskuma* ‘believe’, constitute activities or states in the Aktionsart system of Vendler (1957), which are lexically atelic as they lack inherent temporal end points.

However, Estonian partitive also remains multifunctional as it retains its older partitive function (*kolm õuna* three.NOM apple.PAR ‘three apples’) in addition to its structural function (*taha-n õuna* want-PRS.1SG apple.PAR ‘I want an apple’) (Cann

and Miljan, 2012, both argument structure and nominal usage are analysed as essentially semantic in). See chapter 3 for a more thorough description of differential argument marking in Estonian.

Interestingly, both the separative and partitive function are also marked by Modern Estonian relative, a separative case (‘out of’), as seen in (29a)–(29b).

- (29) a. Isa tul-i Soome-st koju  
 father.NOM come-3SG.PST Finland-ELA home.ILL  
 ‘Father came home from Finland’  
 b. Enamik laste-st õpi-vad hästi  
 most.NOM child.PL-ELA study-3PL.PRS well  
 ‘Most of the children learn well’

In summary, separative function can develop into bleached structural argument marking function via categories such as partitive where metaphorical spatial meaning is missed by speakers, who interpret it as already bleached, marking nothing but an abstract link between concepts.<sup>6</sup>

### 2.3.3 Economy in speech

Why is this type of grammatical change happening? All the described steps, including lexical items becoming adpositions and semantic cases becoming structural cases, are analysed by Lestrade (2010) as rooted in the linguistic laws pertaining to the economy of speech described by Zipf (1949).

Zipf (1949) describes two forces working in opposite directions — the Force of Unification and the Force of Diversification. The former expresses speakers’ economy, in that the narrower the set of lexical items and the broader the semantic applications of each item, the more economical it is for the speaker to retain and use these items. The latter, however, expresses the hearer’s economy, because a larger set of lexical items with fewer meanings per each item results in more effective processing for the hearer. Hence, the Force of Unification compresses several different meanings onto a single lexical item, which leads to greater abstraction.

Hence, on the one hand, meaning becomes concentrated into cases as the Force of Unification aims to reduce the number of linguistic signs in a language. The Force of Diversification, however, which divides separate meanings between a wider set of lexical items (Zipf, 1949, 19–21), leading to more specific semantic meaning in each item. This is likely the dynamic behind a language incorporating a number of distinct argument structures, sensitive to the lexical semantics of

---

<sup>6</sup> Another connection between separative semantics and event participation is more directly linked to the semantics of the source. For instance, certain verbs of negative emotion in Russian (*izbegat* ‘avoid’), which entail separative semantics, may mark their arguments non-canonically in genitive, which is historically linked to an Indo-European separative case (Nichols, 1983). In addition, the diachronic link between ablative (‘from on top of’) and partitive has also been observed in French, German, Bulgarian, Lezgian, Krongo, Finnish and Basque (Heine and Kuteva, 2002).

verbs. The two forces together maintain vocabulary balance, as is evident from the most famous of Zipf's observations, i.e. the connection between the rank of a word in a frequency table and its frequency value (Zipf, 1949).

In addition, the change is also affected by the Law of Meaning Distribution, which posits that the more frequent the word, the more likely it is to have multiple meanings (Zipf, 1949, 27). Similarly to vocabulary balance, this relationship is systematic as well, as the number of meanings increases as a square root of the frequency of the word (Zipf, 1949, 28). Words with more meanings are used in a larger number of contexts, because they apply to more material (Lestrade, 2010). Lestrade (2010) argues that it must be necessarily frequency which results in a more general meaning, rather than more general meaning causing lexemes to be more frequent. because in the latter instance meanings would not be expected to generalise further when the item becomes more frequent. However, frequent items continue to develop more general functions (Barðdal, 2008; Lestrade, 2010).

Finally, Zipf also showed that more frequent words are phonologically shorter than less frequent words (Zipf, 1949, 65). Increased frequency and phonological reduction are both among the defining aspects of grammaticalisation. As items grammaticalise they become shorter and reduce the effort on the part of speakers. Some suggest that as more frequent items are more predictable, they may not require such precise enunciation as less predictable and less frequent items, leading to phonological erosion over time (Ernestus et al., 2002; Haspelmath, 2008; Lestrade, 2010).

### 2.3.4 Summary

This section argued that the degree of grammaticalisation of an individual case is relevant to describing the system of argument structures in a language. Importantly, it demonstrated that although cases have been shown to have the potential to be highly multifunctional (Västi, 2011; Sirola-Belliard, 2011), this does not entirely contradict with the notion of semantic and structural cases (Wunderlich and Lakämper, 2001; Blake, 2001; Butt, 2006, such as those in). However, the distinction ought not to be viewed as categorical, but as gradual. This idea was already represented in the foundational work of Kuryłowicz (1964), who proposed that such labels as 'structural' and 'semantic' only describe the primary functions of cases (see section 2.3). "[...] all case-forms share both kinds of functions, but the syntactical function is *primary* with "grammatical" and *secondary* with "concrete" cases [...]" (Kuryłowicz, 1964, 179). Such a conceptualisation turns case labels into a proportion issue. If structural uses make up a minority of all uses of semantic cases, their semantic status remains unproblematic. However, as semantic cases are used increasingly in structural functions, identifying them as semantic becomes more problematic.

Observing mixed functionality in a case system raises questions about the nature of argument status itself. If both the bleaching of functions and the gram-

matisation of cases are gradual processes, then argumenthood, too, may be a scalar property. The following section will elaborate on somewhat and entirely gradient approaches to argumenthood.

## 2.4 Gradience in grammar

The topic of this thesis — i.e. oblique argument structures — has frequently come up in the discussions about what may or may not constitute arguments. As many types of arguments, including oblique arguments, have been somewhat difficult to categorise by merely following traditional criteria (see section 2.2.2), some authors have suggested that the notion of argumenthood may have different grades. For instance, rather than classifying elements as arguments or adjuncts, one may instead classify oblique arguments as quasi-arguments (Schütze, 1995; Van Valin and LaPolla, 1997; Koenig et al., 2003).

However, gradient phenomena have not been part of traditional grammatical descriptions, as most traditional frameworks of grammar view most linguistic concepts as discrete categories. Such a categorical approach originally stems from the Aristotelian tradition, where entities are defined based on necessary and sufficient conditions (Robins, 1990). In such approaches, membership of a category is binary and all category members share an equal status.

An opposite approach is taken in frameworks where certain phenomena are viewed as inherently gradient, either as varying on an interval scale (Matthews, 1981; Somers, 1984) or an entirely gradual scale (Arka, 2014). Such theories argue that because gradience is evident from the data on a variety of phenomena, description and theory should accommodate this observation to remain empirically relevant.

The categorical tradition has a much longer history than gradient approaches (see a comprehensive review in Aarts, 2007). Aarts notes that the strong human propensity to see the world as being structured in terms of discrete entities stems from the fact that gradience intuitively appears to introduce vagueness, which is difficult to accept as a natural part of conceptualising the world. This especially applies to science, where the aim is to eliminate vagueness, which is seen as analogous to doubt and lack of knowledge (Aarts, 2007, 12). In this spirit, post-Bloomfieldian structuralists, including Joos (1950) and Hockett (1955) viewed language as being inherently made up of discrete entities with no inherently gradient phenomena. “[...] if we find continuous-scale contrasts in the vicinity of what we are sure is language, we exclude them from language” (Hockett, 1955, 17).

Gradient approaches are newer, going back almost a century (Curme, 1935). However, they did not begin to grow in popularity until the 1960s. Early attempts to “interpret the vagueness [...] in a non-vague manner” (Danesh, 1966, 12) go back to the Prague school in the 1960s (Neustupny, 1966; Zadeh, 1965). The fuzzy set theory in (Zadeh, 1965), for instance, depicted membership in a class as a matter of degree, inspiring Lakoff’s notion of fuzzy grammar (Lakoff, 1973).

Other early proponents of inherently gradient systems include Bolinger (1961), who reacted to the categoricism in the structuralist movement (Joos, 1950; Hockett, 1955) by claiming that continuous phenomena ought to be the object of linguistic study, because they occur in both language design and language use. “When one stops talking about switches and begins to talk about potentiometers, one does not necessarily cease talking about electrical systems” (Bolinger, 1961, 10). One of the most famous early gradient approaches, however, is that of Ross (1972), who developed the notion of a ‘squish’, referring to a set of categories, the border between which are ‘squishy, possibly even quantifiable’ (Ross, 1972, 316).

Summarising the reception of gradient theories, Harris (1993, 220) notes that “Squishiness was not a hit.” One difficulty they pose for description has to do with their lower level of formalisation. For instance, much of the feedback for Ross’s squishy category boundaries focused on the fact that the data Ross observed might sound credible when presented in a simplistic manner, but it fails to provide a clear non-discrete grammar framework (Harris, 1993).

All in all, gradient approaches have largely been rejected for various reasons. Some approaches do not agree that language design inherently involves continuous phenomena. Others think that continuous phenomena exist, but that theoretical frameworks should not reflect their continuous nature. The first position addresses the nature of language while the second is concerned with methodology. While truly gradient approaches would answer ‘yes’ to both of these questions, discrete approaches vary between hard and soft categoricism.

Hard categoricism posits that grammar is entirely gradience-free. “Fuzziness is not present in Grammar in any way,... Rather, fuzziness is in the web, the background knowledge on which language is woven, and therefore it has no effect on the form and function of language” (Bouchard, 1995, 33).

Soft categoricism, on the other hand, posits that grammars might well be gradient, but linguists should avoid formalising gradience. This view is more predominant than hard categoricism, being held by many generative and structural theoreticians. Chomsky has continuously stressed the importance of idealisation, starting from ideal speaker-listeners in an idealised, completely homogenous speech community (Chomsky, 1965, 3), ending with ideal data and concepts (Chomsky, 1995, 7). According to his view, all branches of science require a certain degree of idealisation, because categories help make sense of the large amount of facts one faces. “[...] it is the abstract systems that you are constructing that are really the truth; the array of phenomena is some distortion of the truth because of too many factors, all sorts of things. And so, it often makes good sense to disregard phenomena and search for principles that really seem to give some deep insight into why some of them are that way [...]” (Chomsky, 2002, 99).

However, the divide between gradient and categorical views is not particularly discrete itself. Many gradient views hold that categories exist, even though they have fuzzy borders (Aarts, 2007). In addition, they tend to view categories as

useful chronological starting points. “[...] it methodologically makes sense to adopt an Aristotelian model of the categories as a starting point, which can then progressively be relaxed through falsification, rather than proceeding the other way round” (Aarts, 2007, 224). Hence, linguists are tasked with falsifying the hypothesis of a categorical organisation of language, finding signs of gradience in various linguistic phenomena.

Similarly, most categoricist linguistic theories include gradient phenomena. Examples include generative works discussing gradience in grammaticality (Chomsky, 1965; Belletti and Rizzi, 1988), extraction from relative clauses (Haegeman, 1994), argumenthood (Chomsky, 1981). It also extends to semantic classes (Pinker, 1999, 102), lexical and functional categories (Corver and van Riemsdijk, 2001) as well as word classes (Ackema and Neeleman, 2004).

As the number of linguistic studies using corpus and experimental methods increases, gradience in grammar appears to gain momentum. It appears that while qualitative approaches are able to highlight the potential for gradience in a phenomenon, one ought to use quantitative measures to demonstrate that it systematically applies to a phenomenon as a whole. Hence, a change in the methodological landscape of linguistics is qualitatively expected to affect the way human language is described.

### **2.4.1 Role-based approaches to argumenthood**

One major approach introducing a degree of gradience into argument status is the theory based on Proto-roles. This theory grew out of projectionist theories. Namely, many approaches conceptualise information about arguments as something that pertains to verbs as lexemes (Pesetsky, 1982; Levin and Rappaport-Hovav, 2005; Marantz, 2013). As in such approaches arguments are projected by the verbs’ lexical entries, they are called ‘projectionist’ theories.

The simplest way of conceptualising the lexical information of verbs is to view each verb as including a list of thematic roles in its lexical entry (Fillmore, 1968; Jackendoff, 1972, 1976). Roles such as Agent, Patient, Theme, Instrument, Goal, Source, Experiencer and Stimulus must then be viewed as linguistic primitives, which exist outside of the verb’s meaning (Croft, 1991, 156). While many roles may be specified, only a few of them “are put into perspective” (Fillmore, 1977, 73), i.e. qualify as core arguments — e.g. the buyer and goods for ‘buy’.

However, the simple role list approach has been extensively criticised (Levin and Rappaport-Hovav, 2005). First, precise definitions of roles are nearly impossible to achieve (Schütze, 1995; Schlesinger, 1995). Second, some roles appear to be made up of smaller primitive concepts. For instance, Agents entail volitional, effective, initiative and agentive subroles (Cruse, 1973). After reviewing a number of typological and psycholinguistic studies, Rissman and Majid (2019) conclude that only Agents and Patients can be viewed as being represented by distinct abstract categories, the rest referring to highly heterogeneous sets of elements.

As a solution, generalised thematic roles were proposed, which Dowty (1991) called ‘Proto-roles’ and Croft (1998) ‘super-roles’. Dowty (1991) proposed that there are two Proto-roles — Proto-Agent and Proto-Patient — both of which are cluster concepts, made up of semantic properties or ‘entailments’.

Proto-Agent entailments include:

- volitional involvement in the event or state
- sentience (and/or perception)
- causing an event or change in another participant
- movement relative to the position of another participant
- (exists independently of the event named by the verb)

Proto-Patient entailments include:

- undergoes change of state
- incremental theme
- causally affected by another participant
- stationary relative to movement of another participant
- (does not exist independently of the event, or not at all)

(Dowty, 1991, 572)

In a way, Dowty’s Argument Selection Principle lays out a somewhat formal approach to gradience in argument structure. It posits: “In predicates with grammatical subject and object, the argument for which the predicate entails the greatest number of Proto-Agent properties will be lexicalized as the subject of the predicate; the argument having the greatest number of Proto-Patient entailments will be lexicalized as the direct object” (Dowty, 1991, 576). Hence, if one argument has more Proto-Agent properties than another argument, it will be realised as the subject of the clause. Importantly, this means that various elements in a clause vary in terms of their ‘Agenthood’.

In addition, Proto-roles have been influential in non-canonical argument marking studies. Their entailments cover many of the transitivity factors proposed by Hopper and Thompson (1980). Given the hypothesis that transitivity plays a role in case selection in arguments, these entailments have been used for explaining the variance in non-canonical marking in many languages.

For instance, Lestrade (2010) uses them to explain the entire phenomenon of oblique arguments. He proposes that semantic cases are sometimes used in structural function to explicitly mark the lack of Proto-role properties. In this view, spatial case may be used on the subject to communicate that it is non-volitional (Lestrade, 2010, 187) (e.g. the allative-marked Experiencers in Estonian). This would mean that there are more types of realisation than just something being

realised as a (nominative) subject or an (accusative) object Dowty (1991). We will return to this question in chapter 7.

One issue with generalised roles is that they were designed to describe prototypically transitive verbs where an Agent acts on a Patient. They do not address the wide variation in transitivity that languages of the world exhibit (Levin and Rappaport-Hovav, 2005, 60). However, some have argued that Proto-role entailments also play a role in the argument realisation of ‘syntactically intransitive’ verbs (oblique argument verbs) (Davis and Koenig, 2000). For instance, the realisation of subjects and oblique complements of verbs such as ‘rely on’, ‘yearn for’, ‘apply to’ also depend on Proto-Agent and Proto-Patient entailments (Davis and Koenig, 2000, 74).

Some have also criticised the tendency to regard semantic entailments as having the same weight when they are used for determining which constituents become realised as subjects and objects (Davis and Koenig, 2000; Davis, 2001; Ackerman and Moore, 2001; Murphy, 2002). Others have pointed out that the morphosyntax of arguments can well originate from other sources, rather than the entailments, imposed by verbs on their arguments. For instance, object animacy appears to play a role in the way indirect objects are realised in English (‘Send Mary the goods’, \*‘Send the warehouse the goods’) (Evans, 1997).

In summary, the Proto-role approach describes a system which entails gradience in argumenthood to some degree.<sup>7</sup> This gradience, however, only applies to the semantic level, because once a phrase becomes realised as a subject or object, the idea of gradience is no longer applicable to its subject/object status — it either is or is not an object. There are several frameworks, however, which also integrate gradience into the final expression of argument realisation. The next section will outline several such approaches.

## 2.4.2 Gradient approaches to argumenthood

Argumenthood is phenomenon which has often received somewhat gradient interpretations, even in categoricist approaches. Certain types of constituents have been noted to be difficult to fit in a binary model of arguments and adjuncts. For instance, some type of semi-argumenthood has been imposed on subjects of nominalisations (‘**The enemy’s** destruction of the city’) and ‘by’-phrases in passives (‘The city was destroyed by **the enemy**’) (Grimshaw, 1990, 109), the weather-*it* (‘**It** is raining.’) (Chomsky, 1981, 325) and phrases selected by verbs but not fixed in terms of adpositions or cases (‘Kim put the book **on the box**’/‘**under the table**’) (Van Valin, 2005, 23).

---

<sup>7</sup> Many modern versions of projectionist theories have also decomposed verb meanings rather than roles (Pesetsky, 1982; Levin and Rappaport-Hovav, 2005). Decomposed predicate theories propose a more complex picture where the information in a verb’s lexical entry is rather abstract, being realised as arguments via a set of mapping algorithms. A verb’s lexical entry “registers some kind of semantically anchored argument structure, which in turn determines the morphosyntactic expression — or projection — of its arguments.” (Levin and Rappaport-Hovav, 2005, 186).

Broadly, there are two ways to conceptualise gradience in argumenthood. One can view the distinction between argumenthood and adjuncthood as an interval including predefined intermediate levels (Matthews, 1981; Somers, 1984). Alternatively, one can view argumenthood as a fully gradient phenomenon with various, *numerical* argument strengths<sup>8</sup> (Arka, 2014).

Representing an interval approach, Matthews (1981, 140) proposes that the sentence ‘I saw him clearly yesterday’ includes three different levels of arguments: a complement (‘him’), a non-complement (‘clearly’) and peripheral material (‘yesterday’). Somers (1984) posits that argumenthood and adjuncthood entail six categories, the most argument-like of which are internal complements (‘pave the **way**’), followed by obligatory complements (‘**The man** wrote me a letter’), optional complements (‘Graeme caught **Steve** a salmon’), middles (‘I **gently** pushed the button’), adjuncts (‘I pulled the lever **in order to eject**’) and extraperipherals (‘We waited, **you know**’) (Somers, 1984, 524–526). The scale describes an increase in the degree of restrictions — the higher the level of argumenthood (‘complementhood’ in (Somers, 1984)), the more restricted the selected constituents, either in terms of form or semantics. Hence, the ‘middle’ category is similar to a complement in that its form is rather restricted and predictable from the verb’s meaning, but it resembles adjuncts in that its occurrence is entirely optional.

However, if argument status was shown to be based on semantic and pragmatic effects rather than discrete predetermined classes (e.g. ‘optional complements’ in (Somers, 1984)), a gradual scale might have advantages over interval scales. The more gradient approaches include prototype theory (Rosch et al., 1976; Lakoff, 1987) and canonical typology (Corbett, 2005; Brown et al., 2013). Both operationalise gradience by means of well-defined extremes of the scale, against which any observed element can be compared.

### Prototype theory

Prototype theory proposes that while broad categories exist, the members of these categories vary in their prototypicality, i.e. the degree to which they show similarity to a predefined prototype, commonly the most typical member of the category (Taylor, 1995). The idea of prototypes in language goes back to Jespersen (1924), but it gained popularity after Rosch demonstrated that prototypes play an important role in conceptualising real world objects (Rosch, 1973; Rosch et al., 1976). For instance, when assessing the prototypicality of different versions of chairs, the elements assessed to be more prototypical also had more attributes in common with the other members of that category, and less attributes in common with other contrasting categories (Rosch and Mervis, 1975).

This inspired research claiming that prototypes are applicable to linguistic objects (Bybee and Moder, 1983; Lakoff, 1987). Many linguistic phenomena,

---

<sup>8</sup> The terminology of gradient argument status has not set in usage. This thesis refers to arguments of varying degrees as weak and strong arguments.

including noun denotations (Labov, 1973), grammatical categories and syntactic constructions (Taylor, 1995) have been analysed in terms of prototypes. For instance, the prototypical possessive genitive construction can be thought of as entailing an exclusive human possessor, a concrete possessed element and spatial proximity between the two. Based on this, ‘the dog’s bone’ can be assessed as less prototypical than ‘Lisa’s doll’, as ‘the dog’ does not refer to a human entity (Taylor, 1995, 202).

In terms of argumenthood, the prototypicality of arguments is often mapped onto the gradient notion of transitivity (Hopper and Thompson, 1980) and the entailments of the two Proto-roles proposed by Dowty (1991). In this vein, prototypical arguments are defined as those which maximise transitivity in a clause, commonly occurring with highly transitive verbs. Prototypical subjects are volitional and human, meaning that Experiencers constitute non-prototypical subjects (Malchukov, 2005; Lestrade, 2010).

There is, however, some controversy in terms of conceptualising prototypes. Rosch (1978) warns against using prototypes to predict how categories are processed, or thinking that natural language categories have particularly prototypical members. “To speak of a prototype at all is simply a convenient grammatical fiction; what is really referred to are judgments of degree of prototypicality.” (Rosch, 1978, 40).

Similarly, Lakoff (1999, 391) notes that there are two ways in which prototypes are often misinterpreted. First, people assume that whenever speakers give goodness-of-example judgments (‘How good of an example is this entity of this category?’), the membership of that category must be a matter of degree. The second misconception is that the more similar an entity is to a prototype, the more that entity itself is a member of that category. Hence, Lakoff (1999) emphasises the distinction between prototypicality and gradient membership to a category. Simply put, odd-looking chairs are no less chairs than prototypical chairs and one must not conflate the two concepts.

Furthermore, Aarts (2007) remains critical about the proposal of Lakoff (1987), that because linguistic categories show prototype effects, they must have “the same character as other conceptual categories” (Lakoff, 1987, 67). He notes that it is simpler to have participants assess the typicality of chairs than, for instance, the typicality of adjectives, because the properties underlying the assessment differ fundamentally (Aarts, 2007, 88).

Finally, some approaches assume that prototype effects correlate with frequency. Rosch et al. (1976) list five properties of prototypes: 1) prototypes achieve higher typicality ratings from participants, 2) they are learned earlier than other category members, 3) it takes less time to verify them as category members, 4) their representations are generated upon hearing the category name and 5) they are more likely to be named by participants in task of naming category members (Rosch et al., 1976, 491-492). These properties, however, are also the properties

of frequent lexemes. This has given rise to the idea that prototypes not only are frequent in experimental output (Rosch et al., 1976), but they can also be identified as the most frequent elements in natural language (Stamenković and Tasić, 2013). Taylor (1995, 52) notes that “Interestingly, the impression of a higher frequency of occurrence of prototypical members may well be a symptom of prototypicality, and not its cause.”

All in all, the prototype approach outlines a rather intuitive perspective on the gradient nature of categories, but it does not describe inherently gradient variables. Contrary to the way many understand it, it is not clear that prototype effects describe varying degrees of category membership. In fact, this is the way in which linguists have warned others *not* to use prototypes (Lakoff, 1999). Hence, if argument status showed evidence of varying degrees of class membership, it is not clear that the prototype theory would be equipped to describe it.

### Canonical typology

The canonical typology framework has both similarities and differences with prototype theory. It suggests that, for a linguistic phenomenon, one may describe a ‘canon’ or a ‘canonical ideal’. This ideal, however, is conceptualised rather differently from prototypes (Corbett, 2005, 2007; Brown et al., 2013). While prototypes can be viewed as perceptually salient natural categories, a canonical ideal “constitutes a point of convergence for logically consistent definitions.” (Brown et al., 2013, 12). Hence, canonical typology is not a theory of how language is structured, but merely aims to offer a helpful system for conducting typological work. As such, it aims to provide a solution to the Correspondence Problem, which posits that comparing linguistic categories (e.g. verbal agreement or an inflectional class) cross-linguistically requires a way to establish said categories in a universal manner (Brown et al., 2013).

Similarly to some studies in prototype theory, canonical typology describes the canon as defined by multiple dimensions or features (Brown et al., 2013). First, a linguistic phenomenon is defined as a ‘base’, i.e. a broad and inclusive definition, which contains enough information that one can investigate whether this phenomenon exists in a language. Next, criteria are outlined, describing the dimensions along which the base varies.

For instance, if verbal agreement is chosen as the base, the phenomenon may be viewed as more canonical if the controller (nominal) has some visible (overt) expression of agreement features (e.g. the Russian gender marker on the pronoun, with which the verb agrees by bearing the same gender marker: *ona pisala* ‘she wrote’) than when it lacks such features (e.g. the lack of gender marker on the first person pronoun in Russian: *ja pisala* ‘I wrote’) (Corbett, 2006).

These criteria ought to be cross-linguistically informative as they are the dimensions which measure out the space defined by the base. Each of them has a more canonical and a less canonical value. This means that there are various

types of verbal agreement, distinguished by the various configurations of criteria values. In other words, there is only one configuration in which a variant is canonical — in the point where all criteria are valued as canonical — but there are numerous ways in which a phenomenon may be non-canonical. One may also outline the degree to which something is non-canonical, as a quantitative summary of the criteria settings. As more than two criteria are commonly involved, two types of verbal agreement may be non-canonical to the same degree and still express different features.

This approach has also been used for describing argumenthood. Arka (2014) assigned various constituents a numerical argumenthood index, ranging from 0 to 1, reflecting the proportion of 14 different criteria describing the degree to which the argumenthood of each element is canonical. Forker (2014), on the other hand, only postulates six criteria that constitute “defining properties that concern the nature of what it means to be an argument or adjunct” (Forker, 2014, 28). These are:

- Semantic obligatoriness: elements semantically required to complete the meaning of the predicate (see section 2.2.2).
- Syntactic obligatoriness: elements which may not be left out of the clause (see section 2.2.2).
- Latency: implicit arguments require a definite interpretation, available somewhere else in context, while implicit adjuncts do not. One may only understand the meaning of ‘John won.’ if the object of winning is available in context, but the same does not apply to the time or manner of the event.
- Co-occurrence restrictions: in Forker (2014) this criterion entails morphological and semantic selection, as well as the fact that the roles filled by arguments are specific to some verbs (e.g. something being eaten) while adjunct roles are general (the time of the event) (see section 2.2.2).
- Grammatical relations: this pertains to languages that have grammaticalised grammatical relations, i.e. a productive argument structure for marking a wide range of participants, which may be referred to as subjects and objects.
- Iterability: the number of arguments is limited while adjuncts may be added freely (see section 2.2.1).

Forker (2014) uses these criteria to assess the strength of argumenthood in six types of constituents expressing six different functions of two spatial cases in Hinuq (Northeast Caucasian, Russia). Certain functions of these cases are shown to resemble canonical adjuncts (spatial, temporal, manner functions) while others resemble canonical arguments (possessors, non-canonical agents). Hence, differently from prototype theory, canonical typology relies less on transitivity factors and Proto-role entailments and more on traditional argument criteria described (see section 2.2).

### 2.4.3 Summary

In a way, gradient approaches to argumenthood show a great deal of similarity to one another. They mostly assume that variance in argumenthood depends on various predetermined semantic, morphological and syntactic factors. Following this, they also assume that oblique arguments have decreased or inherently different argument status compared to that of canonical arguments, as per the mentioned factors. In this sense, they all pattern with the traditional grammatical description in which obliques are viewed as qualitatively distinct from canonical arguments.

In addition, the two most gradient views on argument status, prototype theory and canonical typology, are not really equipped to measure the degree of membership of an element in a category. Canonical typology explicitly denounces such an aim, outlining itself merely as a tool for typological investigations. Hence, the criteria in canonical typology are designed only so they may capture cross-linguistic variation in a phenomenon, rather than constitute actual cognitive factors affecting the phenomenon in language or describing degrees of membership to classes. This is not always done in studies using canonical typology, though. Several still assess gradience in membership to argument class, building it on traditional argument status criteria, some of which were shown to be highly problematic in section 2.2 (Arka, 2014; Forker, 2014).

The development and popularisation of gradient approaches accompanies the widening spectrum of methods. As quantitative corpus and behavioural methods become more widespread, gradient representations may become more widespread in describing grammar. This means one needs to be careful in terms of interpreting results, because there is still very little knowledge on the role of gradience in grammar. One might need to critically assess whether variance found in quantitative studies really constitutes evidence of gradience in a phenomenon. While empirical (quantitative) methods are undoubtedly more efficient in mapping the scope of potential gradience in a phenomenon than more traditional example-based studies, the final assessment will nevertheless require a combination of data and theory to validate the idea of gradience in a grammar domain (Aigro, 2018).

In any case, a gradient approach would need to possess the ability to describe varying degrees of membership in a class. That is, it would need to have the terminology with which to outline weaker and stronger arguments, as well as the terminology with which to do it. One may argue whether the aspect of varying degrees of membership essentially contradicts the entire concept of classes. For instance, one does not postulate classes for describing human height, but rather defines a type of scale. These issues will be addressed in chapter 6, which uses an experiment to demonstrate the way particular types of elements systematically occupy the argumenthood scale.

### 3 AN OVERVIEW OF ESTONIAN

Estonian is a Finno-Ugric language from the Uralic language family with rich morphology and a varied morphosyntactic system for marking verbal dependants. Together with Finnish, Livonian and several other languages, it belongs to the Finnic branch of Finno-Ugric languages, which are spoken around the Baltic Sea and its neighbouring areas. It is the native language of about 1.1 million people.

As this thesis focuses on arguments marked in spatial cases, the following sections will outline the system of Estonian nominal morphology, focusing on spatial cases and their functions (sections 3.1–3.2). Canonically and non-canonically marked arguments are introduced (sections 3.3–3.5). Section 3.7 presents the research questions of the seven individual studies conducted in this thesis, also discussing its general aims and scope.

#### 3.1 Estonian nominal inflection

The Estonian nominal inflection paradigm describing the inflection of nouns, adjectives, numerals and pronouns includes 28 cells with singular and plural forms of 14 cases. In addition, 26 inflection classes have been distinguished for nominals (Viks, 1992). From a broad typological perspective, Estonian is an agglutinating language, but it includes more fusional morphology than other languages in the northern branch of Finnic (Viitso, 2003).

The nominal paradigms of *õpetaja* ‘teacher’ and *mõte* ‘thought’ are presented in Table 1:

**Table 1:** Estonian nominal paradigm for *õpetaja* ‘teacher’ and *mõte* ‘thought’, including three structural cases (nominative, genitive and partitive) as well as 11 semantic cases. Hyphenation indicates the case affix for singular forms and the plural and case affix for plural forms.

Case	‘teacher’		‘thought’	
	singular	plural	singular	plural
Nominative	õpetaja	õpetaja-d	mõte	mõtte-d
Genitive	õpetaja	õpetaja-te	mõtte	mõte-te
Partitive	õpetaja-t	õpetaja-id	mõte-t	mõtte-id
Illative (‘into’)	õpetaja-sse	õpetaja-te-sse	mõtte-sse	mõte-te-sse
Inessive (‘in’)	õpetaja-s	õpetaja-te-s	mõtte-s	mõte-te-s
Elative (‘out of’)	õpetaja-st	õpetaja-te-st	mõtte-st	mõte-te-st
Allative (‘onto’)	õpetaja-le	õpetaja-te-le	mõtte-le	mõte-te-le
Adessive (‘on’)	õpetaja-l	õpetaja-te-l	mõtte-l	mõte-te-l
Ablative (‘from on top of’)	õpetaja-lt	õpetaja-te-lt	mõtte-lt	mõte-te-lt
Translative (‘become’)	õpetaja-ks	õpetaja-te-ks	mõtte-ks	mõte-te-ks
Terminative (‘up to’)	õpetaja-ni	õpetaja-te-ni	mõtte-ni	mõte-te-ni
Essive (‘as’)	õpetaja-na	õpetaja-te-na	mõtte-na	mõte-te-na
Abessive (‘without’)	õpetaja-ta	õpetaja-te-ta	mõtte-ta	mõte-te-ta
Comitative (‘with’)	õpetaja-ga	õpetaja-te-ga	mõtte-ga	mõte-te-ga

The nominal paradigm includes three structural cases — nominative, genitive and partitive, the term ‘structural’ (or ‘grammatical’) referring to their function as markers of grammatical subjects and objects (grammatical relations). Nominative is morphologically unmarked, but genitive and some instances of partitive are marked fusionally in the stem (see genitive singular in Table 1). The three structural cases may also remain unmarked altogether (*ema* is the nominative, genitive and partitive singular form of ‘mother’).

In addition to the three structural cases, the nominal paradigm includes eleven semantic cases, one of which (illative singular) is sometimes fusionally marked in the stem (*maja* house.NOM, *majja* house.ILL). The other ten are formed by adding a case affix to a stem — the stem often being the genitive form in their respective number (Viitso, 2003; Viht and Habicht, 2019). Six of the semantic cases are regarded as spatial cases: illative, inessive, elative, allative, adessive and ablat-

ive (see Table 1). These cases are further discussed in section 3.2). Sometimes terminative (‘up to’) is viewed as the seventh spatial case (Ylikoski, 2011).

Finally, the last four cases in Table 1 (terminative, essive, comitative and ablative) behave somewhat differently from the rest as they exhibit clitic-like rather than affix-like behaviour. For instance, in these cases adjectives do not agree with head nouns in case or number, but occur in genitive instead (*kollase maja-ni* yellow.GEN house-TERM ‘up to the yellow house’, \**kollase-ni maja-ni* yellow-TERM house-TERM) (Viitso, 2003; Erelt et al., 2007; Miljan, 2009). In addition, the affixes of these four cases attach to phrases rather than nouns (*kassi ja koera-ga* cat.GEN and dog-COM ‘with the cat and dog’, ?*kassi-ga ja koera-ga* cat-COM and dog-COM).

In addition, the nominal paradigm entails a high degree of overabundance, i.e. acceptable parallel forms in cells (Siiman, 2019; Viht and Habicht, 2019). This phenomenon affects most plural cells (e.g. *jooksja-te-sse* and *jooksja-i-sse* runner-PL-ILL), but parallel forms also occur in illative singular (e.g. *maja-sse* house-ILL and *majja* house.ILL) (Siiman, 2019).

Overabundance in the Estonian nominal paradigm broadly arises in three different ways. First, a nominal may have parallel forms in a single paradigm, commonly regarded as a property of particular inflection classes (e.g. class no 25 of the 26 inflection classes in Estonian has parallel forms in most plural cells, such as *harilike-sse* and *harilikkude-sse* ordinary.PL-ILL ‘into the ordinary’). Second, a particular lexeme may have parallel paradigms, both belonging to the same inflection class (*heeringa-te-sse* and *heeringa-i-sse* herring-PL-ILL). Third, a nominal may simultaneously feature in several inflection classes (*akvaariumi-de-sse* and *akvaariume-i-sse* aquarium-PL-ILL) (Ülle Viks, personal communication). According to Kaalep (2018, 146), 13 of the 26 inflection classes have systematic overabundance throughout most of the plural cells of eleven semantic cases while another seven classes feature some overabundant lexemes in those cells.

### 3.2 Spatial cases

The six spatial cases form a coherent semantic system where three cases refer to external space and three to internal space. In addition, three basic direction categories, i.e. lative, locative and separative, are represented in each group, as illustrated in Table 2:

**Table 2:** Estonian spatial cases

	Lative (Goal)	Locative (Place)	Separative (Source)
External (ON)	Allative ( <i>-le</i> , ‘onto’)	Adessive ( <i>-l</i> , ‘on’)	Ablative ( <i>-lt</i> , ‘from on’)
Internal (IN)	Illative ( <i>-sse</i> , ‘into’)	Inessive ( <i>-s</i> , ‘in’)	Elative ( <i>-st</i> , ‘out of’)

In a diachronic perspective, the configuration with three distinct directions already occurred in Proto-Uralic. Of the original three spatial affixes, two have evolved into other cases in Estonian. The locative affix *-na* evolved into essive and the separative *-ta* evolved into partitive (history of partitive was discussed in section 2.3.2). Nevertheless, the three-way system remained in place (Rätsep, 1979) and other forms now fill those functions.

The distinction between internal and external cases is newer, only dating back to Proto-Finnic (Rätsep, 1979). Estonian internal cases are formed by the historical Finnic lative case affix *-s* with an additional element being added in illative (*-sse*) and elative (*-st*).

External cases are formed by the Finnic affix *-l*, which is combined with additional elements in allative (*-le*) and ablative (*-lt*). Allative is likely the oldest external case, dating back further than Proto-Finnic. The other two external cases, adessive and ablative, may have come to use via analogy with internal cases (Rätsep, 1979, 52).

Ross (1997) notes that the only spatial case present in most 17th century texts is elative, while most grammars between the 17th and 19th centuries either only include elative or refer to both elative and allative. Elative is often linked to or presented as a subclass of genitive in these grammars, while allative, when present, occurs under dative (Ross, 1997, 186).

### 3.2.1 Morphology

Morphologically, all six spatial cases are generally marked by affixes added to the genitive form. Illative, however, is slightly distinct from the other five cases. The illative singular cell often has overabundance in the sense that it is expressed by two parallel forms, sometimes called the short (often fusional) illative (30a) and long (agglutinative) illative (30b) (Siiman, 2019):

- (30) a. Lektor astu-s auditooriumi  
 lecturer.NOM step-3SG.PST auditorium.ILL  
 ‘The lecturer stepped into the auditorium’  
 b. Lektor astu-s auditooriumi-sse  
 lecturer.NOM step-3SG.PST auditorium-ILL  
 ‘The lecturer stepped into the auditorium’

Long illative (30b) is formed by adding *-sse* to the genitive stem, while short illative can be expressed by a number of different means, e.g. the bare genitive stem (30a), gemination of stem phonemes (*jõgi* river.NOM *jõkke* river.ILL ‘into the river’), *-he/-hu/-ha*-final fusional forms (*pea* head.NOM *pähe* head.ILL ‘into one’s head’) or with the *-de* suffix (*suur-de* big-ILL ‘into a big’). This means that illative is the only semantic case which can sometimes be marked with a fusional affix.

In lexemes for which the two forms are in parallel distribution, the short form tends to dominate (Viht and Habicht, 2019, 296). However, entirely parallel distri-

bution is rare as lexical and clause factors tend to guide the choice between forms. In a corpus of 15 million words representing written Estonian, Aigro and Vihman (in review) found that it is only for 5% of lexemes with an illative singular form that two parallel forms are simultaneously in use.

Metslang (2014) notes that the choice between two illative forms depends on inflection class, lexeme-final phoneme and lexeme length in syllables. Siiman (2015) argues that the choice of form also depends on whether the lexeme occurs in argument function, but this was shown to not apply for lexemes where both parallel forms of illative singular (or parallel forms in the plural cells of other spatial cases) are used (Aigro and Vihman, in review). However, the finding that one form dominates in a more semantic function while the other form dominates in argument function is consistent with studies in Czech where one variant of locative singular has been found to be preferred over the other in governed position (Bremel and Knittl, 2012).

The distinctions in morphology, semantics and lexical patterns have given rise to the idea that instead of a single illative, two different cases are observed: additive in (30a) and illative in (30b) (Viks, 1982). However, this view has not been accepted by mainstream descriptions (Viht and Habicht, 2019), which emphasise that the two illative forms may well be in parallel distribution and that they are not semantically distinct (Erelt et al., 1995, 56). In addition, both forms are not available to all lexemes.

### 3.2.2 Functions

Most studies approach Estonian spatial cases from a functional viewpoint (Matsumura, 1994; Nurka, 2014; Klavan, 2012; Lindström and Vihman, 2017). The functional perspective relates to section 2.3, which discusses case functionality from the perspective of grammaticalisation. Semantic cases vary in terms of their functional range. The wider the extent of this range and the more bleached functions a case is able to express, the more grammaticalised it is. This thesis regards marking core arguments as the most highly bleached function of a case. The structural use of Estonian semantic cases is discussed in section 3.5. The following will outline the more semantic functions of Estonian spatial cases.

The most straightforward function of spatial cases is that of marking relations pertaining to physical space, as in (30). Concrete spatial meaning is often extended into the function ABSTRACT SPACE (*hinge* soul.ILL ‘into one’s soul’). However, spatial cases have also been noted to have various other functions, which are metaphorically related to the notion of space. Section 2.3 discussed the pathways leading from spatial functions to more bleached functions (Heine, 1990), passing through TIME, PURPOSE and BENEFACTIVE, among others. Most of these are observed in the use of Estonian spatial cases as well. The following outlines the functions of Estonian spatial cases, based on a reference grammar (Erelt et al., 2007).

## Internal cases

The reference grammar (Erelt et al., 2007) outlines five different functions for illative, six for inessive and 13 for elative. All three cases are said to encode 1) SPACE, 2) STATE-OF-AFFAIRS<sup>9</sup>, 3) TIME and 4) bleached function.<sup>10</sup>

More individual functions include REASON for illative (31a) as well as lexic-alised expressions and DOMAIN for inessive (31b):

- (31) a. Ta suri vähki  
3SG.NOM die.3SG.PST cancer.ILL  
'S/he died of cancer'
- b. Ületa-n teisi tennise-s  
surpass-1SG.PRS others.PAR tennis-INE  
'I surpass others in tennis'

Elative, however, also marks COMPARISON (32a), MATERIAL (32b) and the Source in change verbs (32c):

- (32) a. suurem päikese-st  
bigger sun-ELA  
'Bigger than the sun'
- b. betooni-st maja  
concrete-ELA house.NOM  
'A concrete house'
- c. minu-st saa-b õpetaja  
1SG-ELA become-3SG.PRS teacher.NOM  
'I will become a teacher'

In addition, elative marks REASON (33a), the semantic subject in passives (analogous to REASON) (33b), an attribute (analogous to MATERIAL) (33c) and something with which connection is cut (analogous to ABSTRACT SPACE) (33d):

- (33) a. Sõrm-ed olid külma-st kange-d  
finger-NOM.PL.NOM COP.3PL.PST cold-ELA stiff-PL  
'Fingers were stiff from the cold'
- b. Naine oli kuumuse-st piina-tud  
woman.NOM COP.3SG.PST heat-ELA torture-PPP  
'The woman was tortured by the heat'
- c. Üliõpilase-st poeg  
student-ELA son.NOM  
'A student son'/'A son who is a student'
- d. Ta jäi raha-st ilma  
3SG.NOM remain.3SG.PST money-ELA post  
'S/he was left without money'

<sup>9</sup> Expressed as *seisund* 'state', but really entails examples of abstract space, e.g. *vaju-s mõtte-sse* fall-PST.3SG thought-ILL 'fell deep in thought'

<sup>10</sup> Expressed as "Person, thing or phenomenon which the event entails", "function without meaning" or "the object at which the event is aimed" (Erelt et al., 2007).

Finally, elative also marks PARTITIVE function, which in some contexts leads it to be in parallel distribution with partitive (34):

- (34) a. Maarja sõi pool leiba  
 Maarja.NOM eat.3SG.PST half.NOM bread.PAR  
 ‘Maarja ate half the bread’  
 b. Maarja sõi pool leiva-st  
 Maarja.NOM eat.3SG.PST half bread-ELA  
 ‘Maarja ate half the bread’

The elative affix has stood out in terms of its multifunctional usage for more than a century (Wiedemann, 1875; Saluveer, 1958; Viht and Habicht, 2019). Wiedemann (1875) refers to twelve different functions of elative, including SEPARATIVE, MANNER, ORIGIN, PARTITIVE, TIME, and others.

In an especially elaborate approach, Nurka (2014) described elative by means of a multidimensional matrix of 18 different senses in 6 cognitive domains (time, space, abstract space, abstract relationships, containment and scale). Out of the 108 possible functions created by this matrix, she found that elative tokens were used in 48 (sample size = 500 elative tokens). Among the 18 senses of each domain in this matrix, there are three structural functions. These will be further discussed in section 3.5.

### External cases

The reference grammar (Erelt et al., 2007) outlines eight different functions for allative, seven for adessive and three for ablative. All three external cases are used for marking SPACE. Allative and adessive both also mark STATE-OF-AFFAIRS<sup>11</sup> and TIME.

Additional allative functions include RECIPIENT and EXPERIENCER. Allative is also said to have three separate bleached functions, the first referring to “someone or something at which the event is aimed”, the second referring to “the object of an emotion” while the third is “without a clear meaning” (Erelt et al., 2007). Broadly all three constitute bleached functions where allative is selected by the verb, i.e. structural function. These will be further discussed in section 3.5.

In addition to the aforementioned senses, the adessive functions also include POSSESSOR (35a), semantic subjects of infinitives (35b), INSTRUMENT (35c) and MANNER (35d).

- (35) a. Mu-l on koer  
 1SG-ADE COP.3SG.PRS dog.NOM  
 ‘I have a dog’  
 b. Luba neil rääki-da  
 allow.IMP 3PL.ADE talk-INF1  
 ‘Let them speak!’

<sup>11</sup> Similarly to internal cases, this function marks instances of abstract space

- c. Monika mäangi-b klaveri-l lugu  
 Monika.NOM play-3SG.PRS piano-ADE song.PAR  
 ‘Monika is playing a song on the piano’
- d. Ema kuula-b kikkis kõrvu-l  
 mother.NOM listen-3SG.PRS stiff ear.PL-ADE  
 ‘Mum was listening hard’/‘Mum was all ears’

In addition to SPACE, ablative also marks SOURCE, a human role opposite to Recipients, e.g. the SOURCE of *saama* ‘receive’. Finally, ablative marks ATTRIBUTE (*hinge-lt noor* soul-ABL young ‘young by one’s soul’).

Table 3 lists all functions coded in this study:

**Table 3:** Functions of Estonian spatial cases in *Eesti keele käsiraamat* (Erelt et al., 2007).

Illative ‘into’	Inessive ‘in’	Elative ‘out of’
SPACE	SPACE	SPACE
STATE	STATE	STATE
TIME	TIME	TIME
BLEACHED FUNCTION	BLEACHED FUNCTION	BLEACHED FUNCTION
REASON	LEXICALISED PHRASE	COMPARISON
		MATERIAL
		SOURCE
		SEMANTIC SUBJECT IN PASSIVES
		ATTRIBUTE
		SOMETHING WITH WHICH CONNECTION IS CUT
		PARTITIVE
Allative ‘onto’	Adessive ‘on’	Ablative ‘from on’
SPACE	SPACE	SPACE
STATE	STATE	SOURCE
TIME	TIME	ATTRIBUTE
EXPERIENCER	POSSESSOR	
RECIPIENT	SEMANTIC SUBJECT IN INFINITIVES	
SOMETHING AT WHICH THE EVENT IS AIMED	INSTRUMENT	
OBJECT OF EMOTION	MANNER	
BLEACHED FUNCTION		

## Alternations

The extensive multifunctionality of spatial cases leads to their parallel distribution in some contexts. More than a century ago, Wiedemann (1875) pointed out that when marking abstract relations, internal and external cases are sometimes interchangeable (e.g. *hakka-b tö-le* start-PRS.3SG work-ALL and *hakka-b tö-se* start-PRS.3SG work-ILL ‘begins working’). This phenomenon also occurs in Modern Estonian where allative tends to be included in a range of spatial case alternations in argument contexts (Viht and Habicht, 2019, 300). For instance, *sarnanema* ‘resemble’ varies between comitative (36a) and allative structures (36b):

- (36) a. Joosep sarnane-b peaministri-ga  
Joosep.NOM resemble-PRS.3SG prime.minister-COM  
‘Joosep resembles the Prime Minister’  
b. Joosep sarnane-b peaministri-le  
Joosep.NOM resemble-PRS.3SG prime.minister-ALL  
‘Joosep resembles the Prime Minister’

Such parallel uses may also lead to verbs changing their case selection over time. For instance, a century ago Kettunen (1924, 59) mentioned the verb *õnnestuma* ‘succeed in’ as selecting an allative Experiencer, whereas in Modern Estonian it selects an adessive Experiencer. In any case, from the perspective of case evolution, such effects are highly expected as they illustrate the functional similarities between spatial cases (see section 2.3).

### 3.3 Grammatical relations

Typologically, Estonian is a nominative-accusative language, where subjects of both one-place verbs (‘run’) and two-place verbs (‘hit’) occur in nominative. Overall, the set of Estonian grammatical relations are regarded as including six different types of information: 1) the predicate (the verb), 2) grammatical subject, 3) grammatical object, 3) predicative (*koer* ‘dog’ in *Muki on koer* ‘Muki is a dog’) 4) adverbial and 5) attribute (adjectives). This section outlines grammatical subjects and objects, both of which exhibit differential case marking with various semantic and pragmatic implications.

#### 3.3.1 Differential subject marking (DSM)

Grammatical subjects occur in either nominative or partitive. Nominative subjects are prototypical and the verb agrees with them in number and person, e.g. (37a)–(37b). Partitive subjects occur in clauses with decreased transitivity, e.g. in existential clauses (37c), as well as in resultative clauses, with Possessors and Experiencers (Erelt et al., 2017, 244):

- (37) a. Me korja-me seeni  
 1PL.NOM pick-PRS.1PL mushroom.PAR.PL  
 ‘We are picking mushrooms’  
 b. Metsa-s kasva-vad seene-d  
 forest-INE grow-PRS.3PL mushroom-NOM.PL  
 ‘Mushrooms grow in the forest’  
 c. Metsa-s kasva-b seeni  
 forest-INE grow-PRS.3SG mushroom.PAR.PL  
 ‘There are mushrooms growing in the forest’

As the clause in (37a) depicts a transitive event, no differential marking is available. (37b)–(37c), however, depict less transitive existential clauses with available DSM between nominative (37b) and partitive (37c).<sup>12</sup> Note that number agreement only occurs on verbs with nominative subjects (37a)–(37b). When DSM is available, the choice between nominative and partitive is facilitated by various semantic, pragmatic and clause-level factors (Nemvalts, 2000; Metslang, 2012; Huumo and Lindström, 2014), e.g. negation, definiteness and various quantificational properties (Metslang, 2012; Huumo and Lindström, 2014).

### 3.3.2 Differential object marking (DOM)

Estonian grammatical objects may be marked in partitive, genitive or nominative. However, most of them occur in partitive. This applies to all imperfective verbs, which according to Miljan (2009) map on activity and state verbs in (Vendler, 1957), e.g. *vihkama* ‘hate’. However, some verbs (Vendler’s achievement and accomplishment verbs, e.g. *joonistama* ‘draw’) allow alternation between three object cases — partitive, genitive and nominative (Metslang et al., 2017).<sup>13</sup>

In clauses with available DOM, cases convey different aspectual meaning. Partitive marks imperfective aspect (‘partial objects’) while genitive and nominative mark perfective aspect (‘total objects’) (Kont, 1963; Klaas-Lang, 1999; Ogren, 2015; Metslang et al., 2017; Ogren, 2018). This leads to an atelic interpretation for (38a) while (38b)–(38d) describe telic events:

- (38) a. Maarja sõi leiba  
 Maarja.NOM eat.PST.3SG bread.PAR  
 ‘Maarja ate bread’/‘Maarja was eating bread’  
 b. Maarja sõi leiva  
 Maarja.NOM eat.PST.3SG bread.GEN  
 ‘Maarja ate the whole bread’/‘Maarja ate the bread up’  
 c. Söö leib ära  
 eat.IMP.SG bread.NOM up  
 ‘Eat this bread!’

<sup>12</sup> Not all such clauses allow alternation. A number of verbs with existential meanings only select subjects in partitive (*piisama* ‘suffice’) or nominative (*algama* ‘begin’)

<sup>13</sup> The availability of DOM is not only lexical, but also depends on various clause factors, e.g. mood (negated clauses do not have DOM).

- d. Ema            sõi            saiakese-d    ära  
 mother.NOM eat.PST.3SG pastry-NOM.PL up  
 ‘Mum ate all the pastries’/‘Mum ate the pastries up’

Partitive is regarded as expressing unbounded events, which may be expressed in the indefinite quantity of object affected by the event, indefinite time frame for the event, indefinite result. Genitive and nominative mark bounded events, i.e. definite quantity and/or time frame and/or result (Rajandi and Metslang, 1979; Hiietam, 2003; Miljan, 2009). Nominative commonly occurs with impersonal mood, imperative mood (38c) or infinitives, but it also marks plural objects of regular finite verbs (38d). Hence, a semantic distinction exists between partitive and genitive as well as partitive and nominative. The alternation between genitive and nominative depends on nominal morphology (number) and mood.

In addition, some verbs alternate between a canonical object and an oblique argument, as in (39):

- (39) a. Ma            usu-n            sind  
 1SG.NOM believe-PRS.3SG 2SG.PAR  
 ‘I believe you’  
 b. Ma            usu-n            sinu-sse  
 1SG.NOM believe-PRS.3SG 2SG-ILL  
 ‘I believe in you’

From the perspective of Estonian reference grammars, the verb *uskuma* ‘believe’ sometimes selects a grammatical object (39) and sometimes an adverbial (39b). The status of oblique arguments in Estonian is further discussed in section 3.5. However, similarly to DOM, the alternation of partitive and illative conveys an aspectual distinction, but in the opposite way. In (39), the event is somewhat telic, likely applying to one particular conversation. Hence, partitive in (39) marks more telic semantics than illative in (39b), which describes a more permanent state or the nature of the relationship between two participants.

All in all, even though subjects and objects are traditionally viewed as bleached of semantics, Estonian structural cases often mark them together with aspectual meaning. Additional semantics and pragmatics refers to boundedness and aspect of these events, expressing the quantities of objects, timing information and results, among other aspectual properties. Hence, despite their categorisation as ‘structural’, partitive, genitive and nominative are not entirely bleached in structural function (Miljan, 2009).<sup>14</sup>

<sup>14</sup> The variance introduced by DOM has led some authors to believe that genitive and nominative objects actually bear accusative case, the 15th case of Estonian, which is partially syncretic with the genitive and partially with the nominative paradigm (with no distinct affixes of its own) (Ackerman and Moore, 1999; Hiietam, 2003; Lees, 2004). Accusative would also explain the fact that plurality in genitive objects triggers nominative marking (38d). However, the mainstream view holds that Estonian has no accusative and plurality may well trigger differential marking in nominative (Erelt, 2003; Miljan, 2009).

### 3.4 Semantic use of structural cases

The previous section emphasised that even the structural use of structural case is sometimes not entirely semantically bleached. However, structural cases have an even wider functional range, also marking constituents regarded as adverbials. This is not unusual for structural cases in general, e.g. Latin accusative marks both direct objects as well as destination phrases (Blake, 2001, 32). Similarly, the three Estonian structural cases have semantic functions in that they mark various quantity phrases, which reference grammars regard as “adverbials in object case” (Metslang et al., 2017, 275), e.g. (40):

- (40) a. Lõuna maks-ab viis euro-t  
lunch.NOM cost-PRS.3SG five.NOM euro-PAR  
‘Lunch costs five euros’  
b. Vagun kaalu-b mitme-id tonne  
wagon.NOM weigh-PRS.3SG several-PAR.PL ton.PAR.PL  
‘The wagon weighs several tonnes’

The reference grammar regards the phrases *viis euro-t* five euro-PAR ‘five euros’ in (40a) and *mitme-i-d tonne* several-PL-PAR ton.PAR.PL ‘several tonnes’ in (40b) as adverbials, because they have non-bleached semantic content (Metslang et al., 2017, 275) and because they answer questions such as ‘how much?’ and ‘for how long?’ rather than ‘what?’ (Metslang, 2008).

Nevertheless, they are very similar to objects in several regards. For instance, in some instances their case appears to be selected by the verb. Verbs *maksma* ‘cost’ and *kaaluma* ‘weigh’ in (40a)–(40b) only occur with quantity phrases which are marked in partitive. In addition, the price argument is semantically obligatory for the verb *maksma* ‘cost’ to have meaning. Finally, many such ‘adverbials’ have the same type of differential marking as objects, expressing aspect (Metslang et al., 2017):

- (41) a. Malle jooksis ühe kilomeetri  
Malle.NOM run-3SG.PST one.GEN kilometre.GEN  
‘Malle ran a kilometre’  
b. Malle jooksis ühte kilomeetri-t  
Malle.NOM run-PST.3SG one.PAR kilometre-PAR  
‘Malle ran a kilometre’

The case alternation in the quantity phrases in (41) marks the same type of semantic distinction as it does with objects. The genitive in (41a) marks a telic event while the partitive in (41b) implies progressive reading, e.g. ‘Malle was running a one kilometre race (while something else happened)’. For quantity phrases such as (41b), the reference grammar notes that they are “on the border of objects and adverbials” (Metslang et al., 2017, 276).

### 3.5 Structural use of semantic cases

Similarly to the object-like adverbials described in 3.4, another class of constituents often categorised as adverbials (Veismann et al., 2017) also show grammatical relation properties. These constituents are marked in one of the 11 semantic cases, but these cases are used in semantically bleached functions (Torn, 2006). In the Estonian grammatical descriptions they are often referred to as indirect objects (*kaudobjekt/kaudsihitis*) to signify their similarity to canonical objects, as well as bounded adverbial (*sõltuvusmäärus/rektsiooniadverbiaal*) to signify their distinction from canonical objects (Erelt, 2004; Veismann et al., 2017). However, given that the term ‘indirect object’ is cross-linguistically used with another meaning (the third argument of a three-place verb), this thesis will refer to them as oblique arguments.<sup>15</sup>

Oblique arguments are often regarded as semantic arguments which do not participate in syntactic processes pertaining to canonical objects (Erelt, 2004; Veismann et al., 2017). The set includes a wide variety of verbal dependents, occurring in most semantic cases, including all six spatial cases (42). Oblique arguments may appear as the first (and generally animate) argument of two-place verbs (42a), the second argument of two-place verbs (42b), the third argument of three-place verbs, including both more abstract arguments (42c) as well as Recipient arguments (42d):

- (42) a. Mul-le meeldi-b šokolaad  
 1SG-ALL like-PRS.3SG chocolate.NOM  
 ‘I like chocolate’
- b. Peeter armu-s Toomase-sse  
 Peeter.NOM fall.in.love-PST.3SG Toomas-ILL  
 ‘Peeter fell in love with Toomas’
- c. Mees süüdistab-naabri-t mõrva-s  
 man.NOM accuse-PRS.3SG neighbour-PAR murder-INE  
 ‘The man accuses the neighbour in murder’
- d. Ma and-sin lapse-le pulgakommi  
 1.SG.NOM give-PST.1SG child-ALL lollipop.GEN  
 ‘I gave the child a lollipop’

The mainstream position of the reference grammar is that oblique arguments constitute arguments only semantically, while syntactically they are adverbials (Erelt, 2017c; Veismann et al., 2017). ‘Semantic subjects’ and ‘semantic objects’ are terms that refer to expression of a ‘macrorole’ (Erelt et al., 2017, 244), meaning they describe entities with some degree of Proto-role entailments Dowty (1991) (see section 2.4.1). For instance, the semantic subject of a clause might be expressed as “a sentence-initial adverbial expressing Possessor, Experiencer or a participant causing a change in a state” (Erelt et al., 2017, 244). Similarly, a se-

<sup>15</sup> *Obliikva* has also been used by Erelt (2004) to refer to Location or Manner adverbials closely linked to verbs, e.g. the Manner occurring with *käituma* ‘behave’.

mantic object is a constituent marked in semantic case that is “an argument which shows the entity to which the action is directed”<sup>16</sup> (Erelt, 2017c, 71). All in all, if arguments occur in semantic cases, they are excluded from grammatical subjects and objects (see section 3.3) (Metslang et al., 2017; Veismann et al., 2017).

Not everyone agrees this should be the case. During the past century, several linguists have proposed that oblique arguments ought to be included among syntactic roles. The following sections discuss the two main waves of this discussion, which occurred in the late 1950s and 1980s. They addressed the distinction and similarity between oblique and canonical arguments, as well as between oblique arguments and adverbials.

### The first wave

The structuralist view on argument status, according to which morphological case and syntactic processes bear a defining role in terms of arguments and adjuncts, was influentially introduced in Estonian linguistics by Kettunen (1924). Because oblique arguments are marked in semantic cases, they are excluded from the subject and object class, and instead included in the adverbial class. Kettunen (1924) suggests six different adverbial categories, shown in Table 4:

**Table 4:** Adverbial classes from Kettunen (1924).

Adverbial type	Examples
1. General adverbial	Various adverbials expressing the location, time, manner, reason, function and comitative participants of events. <i>Koer on aias.</i> ‘The dog is <b>in the garden.</b> ’
2. Particle adverbial	Particles pertaining to the verb, such as <i>ka</i> ‘also’ <i>võib-olla</i> ‘maybe’. <i>Oskan ka laulda.</i> ‘I can sing, <b>too.</b> ’
3. Object-adverbial	The elative argument of <i>hoolima</i> ‘care about’, the elative argument of <i>jutustama</i> ‘tell a tale about’
4. Dative adverbial	The allative Goal arguments of <i>andma</i> ‘give’, <i>laenama</i> ‘lend’, the adessive Experiencer of <i>luhta minema</i> ‘fail in’
5. Predicate adverbial	The translative phrase in <i>Jaan oli talus sulaseks</i> ‘Jaan was <b>a farm hand</b> in the farm’, the essive modifiers expressing manner in <i>nägin teda haigena</i> ‘I saw him/her <b>as ill</b> ’, the translative argument of <i>valima</i> ‘elect’.
6. Agent adverbial	The elative Agent in <i>jumalast loodud</i> ‘created <b>by God</b> ’, the genitive Agent in <i>ema istutatud</i> ‘planted <b>by mum</b> ’, the adessive Agent in <i>lase hobusel puhata</i> ‘let <b>the horse</b> rest’, the Agent phrase marked by the adposition <i>poolt</i> ‘by’ in passives, impersonals and past participles.

<sup>16</sup> The author’s translation.

Four out of the six adverbial types in Table 4 (3–6) include various argument-like elements: object-adverbials, dative adverbials, predicate adverbials and agent adverbials.

The structuralist Finnish-inspired view of Kettunen (1924) later was opposed by another author, Kure (1959), whose theory was called the ‘logic-based approach’ by some (Saks, 1959; Vääri, 1959; Lehari, 1959) and the semantic approach by others (Rätsep, 1959). Essentially, this approach posits that semantic arguments and grammatical arguments ought to be viewed as elements on the same hierarchical (argument) level. Various other linguistic features, such as case morphology and syntactic alternations, are regarded as not relevant to defining the concept of core arguments. Kure (1959, 47) notes that the (“morphologist”) view of Kettunen overly values the importance of case.<sup>17</sup> He highlights several issues introduced by the structuralist case-based approach:

1. Oblique arguments are analogous to syntactic relations in that they often refer to participants directly involved in the event — e.g. the inessive complement of *kahtlema* ‘doubt’, the allative Recipient of *andma* ‘give’, the ablative Source of *võtma* ‘take’.
2. Oblique arguments are semantically different from adverbials using the same cases. A speaker immediately recognises that allative marks a much more fundamental role with regard to the verb in Recipients than in spatial adverbials.
3. Cases inherently entail various functions. Kettunen’s view (1924) acknowledges this as a property of structural cases, which may mark subjects, objects (and adverbials), but for some reason does not allow semantic cases to be multifunctional and vary between marking objects and adverbials in the same way.

Kure (1959) therefore views syntactic relations as defined by functions and roles. His proposed oblique argument set is rather wide, including direct and indirect state-of-affair participants (e.g. the elative complement of *sõltuma* ‘depend on’ and the allative Recipient of *andma* ‘give’), as well as comitative participants (*kõnni-b lapse-ga* walk-PRS.3SG child-COM ‘walks with the child’), instruments and material (32b). Constituents expressing the manner, time, amount, reason, purpose, condition, intensity or other such types of information pertaining to events are classified as adverbials. The ‘grammatical connection’ between oblique arguments and verbs is much stronger (as is evident from case selection) than the connection between adverbials and verbs (Saks, 1959, 486).

The semantic-functional approach to argumenthood was not received well (Vääri, 1959; Lehari, 1959). Case functions were claimed to be too complicated

---

<sup>17</sup> Kure cites H. Pöld, J. Jõgever and K. A. Hermann as earlier proponents of this approach.

to be the basis of argument status, oftentimes even indistinguishable between oblique arguments and adverbials (Vääri, 1959). Authors feared that one is unable to distinguish between several adverbial functions (time, location) and other, more bleached functions of cases. They also viewed the phenomenon of oblique arguments as a fad — an artificial, Western inspired concept, originating from the indirect object of Indo-European languages, forcibly introduced to Estonian, which as a syntactically distinct Uralic language cannot accommodate it (Mihkla, 1959; Vääri, 1959).<sup>18</sup> Others claimed that instead of morphology or semantics, syntactic connections between lexemes should be the primary level on which the categories of grammatical relations and adverbs should be based (“syntactic connection” here most likely referring to various syntactic operations applying to canonical subjects and objects) (Rätsep, 1959; Kask, 1960).

As a result of this debate, oblique arguments remained in the class of adverbials in Estonian linguistics (Mihkla et al., 1974).

### The second wave

The similarities between canonical and oblique arguments resurfaced as a discussion topic in the end of 1980s. Klaas (1988) revived the proposal for a special syntactic status for oblique arguments.<sup>19</sup> Similarly to Kure (1959), she posits that the class of adverbials is too wide and that some of them ought to have syntactic relation status.

- (43) a. Ta saabu-s linna-st  
 3SG.NOM arrive-PST.3SG TOWN-ELA  
 ‘He arrived from town’  
 b. Ta rääki-s linna-st  
 3SG.NOM talk-PST.3SG TOWN-ELA  
 ‘He talked about the town’  
 (Klaas, 1988, 38)

On one hand, (Klaas, 1988) maintains that oblique arguments such as (43b) are not to be regarded as objects, because they are morphologically and semantically distinct (Klaas, 1988, 38). Nevertheless, they differ from adverbials (43a) in three respects.

First, oblique arguments and adverbials constitute answers to different types of questions. The relative phrase in (43a) answers the question *kust?* ‘from where?’

<sup>18</sup> Mihkla (1959) essentially sided with Kure (1959). In his view, oblique arguments may be seen as a subcategory of syntactic relations, if it is defined in a rather narrow way, i.e. only including participants of events or states-of-affairs (Themes, Experiencer) as well as the Goals and Sources of various interaction verbs (Mihkla, 1959, 174).

<sup>19</sup> She uses the label *indirektne objekt* ‘indirect object’, which is not meant to refer to three-place verbs, but instead to (mostly) two-place verbs selecting oblique arguments (Klaas, 1988, 37). The present study will refer to them as oblique arguments in order to reserve ‘indirect object’ for the third argument of three-place verbs.

while the relative phrase in (43b) answers the question *millest?* ‘about what?’. In addition, when they are substituted with pronominals, the phrase in (43a) becomes *sealt* ‘from there’ while the phrase in (43b) becomes *sellest* ‘about this’. Finally, that case on oblique arguments is selected by the verb, while case on adverbials tends to vary depending on semantics Saks (1959).

Like the proposal of Kure (1959), this view was also deemed too problematic to be integrated into the mainstream grammatical description (Erelt, 1989). The reasons for this are further discussed in the next section.

### 3.6 The current view

A summary of the mainstream position on core argument status in Estonian is as follows (Erelt et al., 1993; Erelt and Metslang, 2017): all grammatical relations (grammatical subjects and objects) are marked in structural cases. Adverbials are mostly marked in the rest of the 11 cases. Some adverbials might surface as semantic (oblique) arguments, but they nevertheless are qualitatively distinct from core arguments (Veismann et al., 2017).

The approach is based solely on case and syntactic processes. Erelt refers to various syntactic properties (‘syntactic functions’) which describe (and define) grammatical relations (Erelt, 1989, 2003, 2004). For instance, subjects trigger verbal agreement in person and number while objects 1) are marked in nominative in passives, 2) are marked in genitive in *-mine* nominalisation, 3) may trigger number agreement on manner adverbials, 4) control reflexives (Erelt, 2004).

Syntactic processes are also used for expressing the similarities between oblique and canonical arguments. For instance, similarly to subjects, some adessive constituents control reflexives and converbs (i.e. they are the subject of both the finite verb and the converb) (44a). In addition, a raising transformation that describes canonical subjects (44b)–(44c) also pertains to some oblique arguments (44d)–(44e). The following examples originate from Erelt (1989, 15–17):

- (44)
- a. Lapse-l hakka-s seis-tes paha  
child-ADE begin-PST.3SG stand-CVB ill  
‘Child got ill while standing’
  - b. Näi-b et ta tule-b siia  
seem-PRS.3SG that 3SG.NOM come-PRS.3SG here  
‘It seems s/he is coming here’
  - c. Ta näi-b siia tule-vat  
3SG.NOM seem-PRS.3SG here come-3EVID  
‘S/he seems to be coming here’
  - d. On targem et ta tule-b siia  
COP.PRS.3SG wiser that 3SG.NOM come-PRS.3SG here  
‘It is wiser that s/he comes here’
  - e. Ta-l on targem siia tulla  
3SG-ADE COP.PRS.3SG wiser here come.INF1  
‘It is wiser for her/him to come here’

As other oblique arguments share even fewer syntactic operations, they are deemed even less argument-like, e.g. the translative complement of *muutuma* ‘become’, allative Recipients and relative topics of verbs and nominals (43b). Such elements are said to share no meaningful similarities with canonical arguments (Erelt, 1989, 17–18).

### 3.6.1 Issues with the current view

Several issues arise when argument status is viewed as a morphosyntactic phenomenon. Several were already discussed in section 2.2.2, but the following briefly outlines them for Estonian.

First, when syntactic processes are regarded as class membership criteria for argumenthood, only a subset of all canonical subjects and objects pass as arguments (as is also mentioned by Erelt, 1989). For instance, only subjects of clauses high on the transitivity scale are acceptable as nominative subjects in passives, as was shown in (19)–(20). Similarly, becoming a genitive subject of a nominalised verb is also not allowed for subjects of states (*\*mäe linna piiramine* mountain.GEN town.GEN surrounding ‘the mountain’s surrounding of the town’). Partitive subjects are also excluded from the set of arguments, then, as they do not trigger verbal agreement in number.

Second, as discussed in chapter 2, most syntactic properties of core arguments are simply derivatives of morphological case, or more precisely, the predominant canonical argument structure. In Estonian, too, the canonical argument structure developed a wider range of syntactic processes than other argument structures (e.g. a structure where complements occur in illative (42b)). Most alternations, such as becoming nominative in passives and genitive in nominalisations therefore only pertain to this one structure, and only to some of its instances.

The fact that syntactic processes only apply to some structures may be interpreted in several ways. One option, described above, is to view syntactic processes as evidence of the structure’s unique ability to mark core arguments (Erelt, 1989, 2004). This interpretation assumes that when speakers encounter various novel verbs, some of which are strongly linked to particular event/state participants, they subconsciously recognise the fact that the encountered verb has a core participant. Then they retrieve the rule that this type of connection is marked by the canonical argument structure, leading them to use this structure with the novel verb. It is important to note that linking argument status exclusively to one particular structure means making the assumption that it is the *coreness* of some event participants that leads speakers to use the structure. It also follows that it is because of widespread *coreness* that the canonical structure is used more often than other structures, and why it becomes involved with various syntactic operations (e.g. passivisation, verbal agreement, etc). Hence, these syntactic processes and argument status are exclusively linked to one another, which means one may use syntactic processes to identify instances of (core) argumenthood.

Another way to interpret the discrepancy in syntactic processes between argument structures is to view processes as a product of structure productivity. In this view, a language incorporates a range of argument structures, which are essentially linguistic patterns of the same kind and argument marking capability, hypothesised to mark the same phenomenon — an argument-verb connection. However, they differ in terms of productivity. Some argument structures are more integrated into syntax than others, they include cases on a higher grammaticalisation degree and they are more productive. These three properties are interlinked, all leading to further changes in one another. Increasing productivity will lead to further case grammaticalisation, which in turn will increase the variety of syntactic pattern alternations for that structure.

In this view, argument structures do not, however, vary in terms of the quality of argument status they mark. An analogy may be drawn with derivational morphology, where a language may have several morphemes for the deverbal derivation of manner adverbs. Even though these morphemes may vary greatly in terms of productivity, one being used on a much wider range of novel formations than the other, both would nevertheless be regarded as forming adverbials of the same basic type. Hence, it is not a problem that the adessive or elative argument structure render arguments that do not share the syntactic behaviour of canonical arguments — this derives from the fact that they constitute different structures and are therefore expected to be morphologically and syntactically distinct.

All in all, these two opposite views — Erelt's view of argument status being defined by morphosyntax (section 3.6) and the alternative view of argument status being semantically based, essentially outline two plausible hypotheses, which can be tested empirically. Although the semantic approach does not have the advantage of the morphosyntactic view, which makes argument status 'visible' in that one may use simple means to categorise elements, experimental methods have been used to assess the degree of argument status by other means than morphology and syntax. In a design not relying on morphology or syntax, one may use these properties as independent variables to test the suggestion that case type and the range of available syntactic processes are linked to the presence or strength of argumenthood (depending on whether one takes a categorical or gradient approach). This task is undertaken in chapter 6.

### **3.7 This thesis**

This thesis investigates the system of argument structures in Estonian. It focuses on one of its subdomains, i.e. any oblique argument structures including spatial cases. It aims to bring focus away from the canonical argument structure and highlight the other functional members of the set of argument structures used in a language, emphasising the system as a whole. By focusing on one single semantic field — cases sometimes used for expressing space — the thesis is also able to take a more thorough approach by including the grammaticalisation per-

spective of cases, joining the diachronic and synchronic dimensions together. This is because functions may be expected to be somewhat analogous between spatial cases, allowing for cohesive analysis of the emergence and usage of structures.

More precisely, the present thesis is concerned with outlining the similarities and distinctions between a number of argument structures, highlighting the division of labour in the system of argument structures. It tackles oblique structures from three different angles, representing the three distinct parts of an argument structure — the verb, the selected case and the nature of the connection between the verb and the argument.

### 3.7.1 Verbs

First, the thesis investigates argument structures' sensitivity to verb semantics in Estonian, as well as the distinct usage patterns of these verbs. We hypothesise that argument structures vary in terms of verb types (Tsunoda, 1981; Onishi, 2001; Malchukov, 2005; de Hoop and de Swart, 2009). Section 2.2.3 outlined two widespread notions describing this distinction. First, non-canonical argument structures are generally assumed to be assigned to verbs with decreased transitivity (Tsunoda, 1981; Nichols, 1983; Blake, 2001; Malchukov, 2005; Lestrade, 2010). Second, these structures are assumed to reflect some degree of the original semantics of the cases they include (Nichols, 1983; Blake, 2001).

The present thesis tests these two hypothesis in a quantitative corpus study. First, chapter 4 attempts to identify the semantic properties facilitating the assignment of canonical and non-canonical argument structures, comparing the lexical semantics of both types of verbs. Second, it tackles the presence of case semantics. Chapter 3 showed that some bleached roles in oblique structures are indeed somewhat spatial (e.g. allative Recipients as human Goals) while others are merely stative and non-spatial (the relative complement of *sõltuma* 'depend on'). Various argument structures are compared to each other in terms of how often case semantics is reflected in the verbs occurring with them. Finally, it investigates the way structures vary in terms of usage by means of a regression study focusing on the effect of various clause factors.

These studies will address rather profound questions in argument structure research. If argument structure assignment was shown to be affected by the lexical semantics of verbs and/or the presence of the case's original semantics, that would contribute to the view that there is no idiosyncratic case selection in argument structures, directly going against generative views on argument structure (Butt, 2008). Idiosyncratic case selection was an idea promoted by the early discussions on quirky case in Icelandic (Zaenen and Maling, 1984). The view expresses that canonical argument structures are assigned via a syntactic rule, while non-canonical argument structures are idiosyncratically (randomly) defined in the verb's lexical entry. This assumption mostly applies to various generative approaches to case, where a structural distinction is assumed between the assignment

of structural and semantic cases, the latter being idiosyncratic and not thematically conditioned. If, for instance, non-canonical argument structures were shown to be systematically sensitive to the lexical semantics of verbs, that would strongly imply that the canonical structure is semantically conditioned as well. This is because in that case, *all verbs* would need to go through semantic vetting upon their introduction into a language.

In any case, to understand the restrictions, the assignment and usage patterns of various argument structures, one needs to quantitatively explore their usage, as well as the two effects hypothesised to pertain to non-canonical argument structures — decreased transitivity in verbs and spatial semantics in case meaning.

### 3.7.2 Cases

The second element in terms of which argument structures vary, is morphological case. The traditional approach of dividing cases into semantic and syntactic categories was shown to be problematic, both cross-linguistically (section 2.3) and even more so in Estonian (sections 3.3 and 3.5). Hence, a more gradient view on case function is needed, investigating their functionality in a more detailed way.

Furthermore, there is very little knowledge of how cases themselves could affect the usage proportions between various structures. Are structures with some cases more often and/or differently used than those with others? If so, why? This thesis joins together the diachronic and the synchronic perspective of Estonian spatial cases. The evolution of individual cases is highly relevant to understanding the argument structures using them as well as the argument structure system as a whole. This is because the degrees of grammaticalisation of individual cases are directly linked to their functionality in the structural domain. Regardless of this rather direct link, we have no empirical overview of how the different degrees of grammaticalisation in semantic cases contribute to their structural behaviour in terms of the assignment of said structures, the usage of verbs selecting them and their effect on the quality of the cognitive link between verbs and their dependants (argument status).

This is where Estonian is especially useful. Differently from Indo-European languages where restricted case systems may render one or two distinct oblique structures, all six Estonian spatial cases have been noted to occur in structural position and bleached function. This allows us to conduct a thorough study with enough data to express variation in the system, without leaving a single semantic domain — i.e. that of space.

### 3.7.3 Argument status

The third hypothesised distinction between different argument structures is the quality of the link between verbs and arguments they mark. Even though the nature of this connection remains obscure, we may simply note that some verbs appear to extend an underlying cognitive connection to certain other constituents,

which appear to refer to the core participants of the event/state described by the verb. Morphosyntactic approaches assume the link is present in canonical structures and absent or weaker in oblique structures, while semantic approaches assume the link to be equally strong in all argument structures.

Given that experimental approaches have the ability to indicate this clausal connection by other means than the morphology and syntax of the structure, we may test this hypothesis by means of an experiment, as is done in chapter 6. The experiment explores whether syntactically more integrated (canonical) structures automatically mark a stronger connection between verbs and arguments than other (oblique) structures. In addition, we may investigate how the argument status marked by oblique structures patterns in terms of adverbials.

### 3.7.4 Subject matter

As mentioned above, Estonian is an excellent language for conducting a comprehensive study on argument structures, because its rich nominal morphology is accompanied by a rich argument structure system. However, the thesis will only investigate structures with spatial cases. As its six spatial cases form a semantically somewhat coherent group, we are able to investigate the phenomenon in a more systematic manner, both synchronically and diachronically.

Another restriction is that the thesis only investigates a phenomenon sometimes called ‘split alternations’ (Malchukov, 2005), ‘split’ referring to the absence or marginality of case alternation. These are verbs which mainly or exclusively select a particular case to mark their arguments and can therefore clearly be said to be linked to one particular argument structure,<sup>20</sup> e.g. the complement of *sõltuma* ‘depend on’ is always marked in relative. While this thesis does not claim that exclusive use of a single case is the definition of argument structure, it does regard this type of selection as a rather simple and unproblematic instance of an argument structure. It is suitable for the purposes of this thesis, because it yields data that are structurally well-defined and internally coherent. It also means that any comparison between oblique and canonical structures would be on a more equal grounds.

### 3.7.5 Questions and hypotheses

This thesis investigates the three broad topics outlined above (verbs, cases and quality of argument status, see sections 3.7.1–3.7.3). They are the basis for the three broad questions:

---

<sup>20</sup> However, verb usage usually entails a wide degree of variation, so when compiling a sample, we can only claim that a verb is *predominantly* used with a particular structure.

1. How do argument structures vary in terms of the lexical semantics and the usage of verbs? (Chapter 4)

Hypotheses: 1) As was discussed in section 2.2, non-canonical argument structures are expected to occur with less transitive verbs, 2) Given the extensive functionality of Estonian spatial cases (section 3.2.2 and 3.5), this thesis expects to find spatial cases bleached of their spatial meaning in most of the verb sample, 3) We expect to find that non-canonical argument structures are used differently from the canonical structure, but also that individual non-canonical structures show distinct usage patterns, some being closer to the canonical structure than others.

2. How do argument structures vary in terms of the degree of grammaticalisation of the cases they include? (Chapter 5)

Hypothesis: The thesis hypothesises that Estonian spatial cases have very different degrees of grammaticalisation. More precisely, it expects to find that elative and allative are more grammaticalised than other spatial cases. This is expected to reflect back on question 1, because the study expects grammaticalisation degrees to be reflected in both the degree to which case usage is bleached in selected position, as well as in how similar certain oblique structures are to the canonical structure in terms of usage.

3. How does argument status quality vary across argument structures and various types of adverbials? (Chapter 6)

Hypothesis: The thesis hypothesises that various argument structures mark argument status of the same quality, regardless of their morphosyntactic distinctions. The thesis explores various types of adverbials, which might help illustrate the gradience of argumenthood by constituting examples of semi-strong argument status.

Each of these three questions is answered by addressing a number of more detailed research questions (RQs), which will be pursued separately. Chapter 4 addresses question 1 and outlines the three specific RQs contributing to it. Chapter 5 pursues question 2, building on three specific RQs, which together outline degrees of case grammaticalisation. Chapter 6 aims to answer question 3 based on two RQs outlined in section 6.2.3.

The thesis incorporates both corpus and behavioural data to answer its research questions. The novel ways in which this thesis operationalises various semantic and functional concepts into categories one may study by means of quantified variables, constitute an important part of its overall contribution. For corpus studies, this includes the segmentation of verb lexical semantics in chapter 4 and the synchronic approach to grammaticalisation in chapter 5. However, as argument status is an inherently cognitive phenomenon, the ‘visibility’ of which in morphosyntax is debatable, behavioural data is required to investigate it (see Klavan and Divjak, 2016, about the advantages of methodological pluralism in linguistic studies).

## 4 VERBS AND ARGUMENT STRUCTURES

This chapter investigates the system of argument structures in Estonian, focusing on the first of the three major variables in terms of which different argument structures vary — verbs (Aigro and Vihman, in press). It addresses the first of the three broad objectives of this thesis, discussed in chapter 1 and, in more detail, section 3.7, to investigate the distinctions between verbs in oblique and canonical argument structures with regard to lexical semantics (section 2.2.3), the degree to which case usage is bleached, as well as the usage patterns of verbs. All in all, the chapter will attempt to answer the following three research questions (RQs):

RQ1: Are oblique argument structures assigned to lexically less transitive verbs than canonical argument structures? If yes, in what way are these verbs less transitive?

RQ2: To what degree can spatial semantics be detected in Estonian verbs occurring in oblique argument structures and how does this compare to canonical object verbs?

RQ3: Does the usage of verbs with oblique argument structures differ from that of canonical argument structures? If yes, in what way?

We expect to find that non-canonical argument structures occur with less transitive verbs (Malchukov, 2005; Lestrade, 2010). We also hypothesise that Estonian spatial cases (sections 3.2.2 and 3.5) are all occasionally used in non-canonical argument structures with meanings bleached of spatial semantics (contrary to the general hypothesis of Blake, 2001), but that this applies to some more than others (Erelt et al., 2007; Nurka, 2014; Lindström and Vihman, 2017; Viht and Habicht, 2019). Finally, the study expects to find that non-canonical argument structures show distinct usage patterns, some being more similar to the usage of canonical structures.

All three variables — lexical transitivity, degree of original case semantics and verb usage — are therefore expected to show both a structure type based (canonical vs oblique) and structure based effect (between spatial case structures). We also expect the results of the studies presented in chapter to bear a close link to those of chapter 5 in that the variance found in the present chapter is expected to be linked to case functionality and individual degrees of grammaticalisation in the next chapter.

Each RQ is answered by means of a separate corpus study. Two of the three corpus studies are concerned with semantic features of verbs, which means they will be based on the same verb lemma lists, extracted from the corpus. The third study compares verb usage, meaning it is based on clause lists.

The corpus methodology used in all three studies is novel in two respects. First, both the quantitative verb and sentence samples are derived non-semantically,

meaning the studies will be able to test various hypothesis and really hone in on semantic distinctions as they are not built into the samples in the first place. In addition, all three studies compare oblique and canonical argument structures. The thesis holds that it is only by knowing what distinguishes between the two groups that allows one to outline the characteristics of oblique structures and the dynamics underneath the entire argument structure system.

## 4.1 Data

This section describes the assembly of the two datasets that form the basis of answering all three RQs listed above. All data is freely available (Aigro, 2022b).

A list of verbs was extracted for both the oblique and canonical group, based on particular criteria. Based on these verbs, a sample of sentences was compiled for both groups. Out of the three studies in this chapter, studies 1 and 2 are based on a single verb dataset, while study 3 is based on a clause-based dataset.

### 4.1.1 The corpus

Oblique and canonical datasets were both extracted from the Balanced Corpus of Estonian, a morphologically parsed 15-million word corpus of written Estonian, equally divided between journalism (daily and weekly newspapers, 1995–2007), science (dissertations and articles from various disciplines, 1995–2006) and fiction (prose and poetry, 1987–2011).

### 4.1.2 Detecting case selection in a corpus

We propose a novel methodology for extracting argument structure data — i.e. by using co-occurrence frequencies between verbs and complements. This might seem like a controversial choice, because the ways in which co-occurrence frequencies have been used in syntax vary a great deal (the concept was criticised as an argument criterium in section 2.2.2). For instance, some use syntactic obligatoriness to define argumenthood (Barbu and Toivonen, 2016) while others regard it as merely one of the factors contributing to prototypical argumenthood (Arka, 2014; Forker, 2014). The role of co-occurrence as an indicator of argument structure is especially complicated in Estonian, given the frequent ellipsis of both canonically and non-canonically marked subjects and objects (Erelt, 2017a; Lindström and Vihman, 2017) (see chapter 3). For an opposite view, see Vaiss (2021) for an approach where co-occurrence frequencies are viewed as lexical properties of verbs linked to transitivity status.

However, when used as part of a sampling methodology, the question is not about whether non-obligatory arguments may be viewed as arguments, but whether verbs frequently co-occurring constituents with particular morphology regularly constitute instances of argument structure. Another important question is, does this extraction method introduce bias — for instance, are verbs with frequently

overt arguments somehow systematically different from verbs, the arguments of which mostly remain implicit.

The answer to the first question is affirmative, as will be shown in 4.1.3. It is a useful data extraction method, because it renders suitable data and because it has the crucial advantage of not being based on verbal lexical semantics. Comparing two types of verbs is only possible with non-semantically selected verbs, because only then can lexical semantics be regarded as an independent variable.

The second question, however, is more complicated. While Estonian has argument ellipsis on all levels, subjects are especially vulnerable to be omitted, both in spoken and written Estonian and both with canonical and non-canonical case marking (Lindström et al., 2009; Lindström and Vihman, 2017). This might mean that verbs with Experiencer arguments, which show certain similarities to canonical subjects (Onishi, 2001) are less represented than other verbs in such a dataset, given the frequent ellipsis of Experiencer arguments (Lindström and Uiboed, 2017). This would most likely affect the proportion of verbs governing allative, the case linked to Experiencer arguments.

While the thesis acknowledges this as a downside of the present method, for the purposes of this study, it does not regard it as a major issue. The only objective of the methodology used in this chapter is to automatically extract verbs with particular structures. Even if the final dataset includes a smaller proportion of Experiencer verbs than it would under a different method, there is no reason to assume that it would affect a major part of the potential sample (Experiencer function only applies to one case, allative, see section 3.2.2). Furthermore, conclusions are not drawn from the proportions of different argument structures found in this chapter (as this is done in chapter 5 by using other methodology). Hence, regardless of this minor disadvantage of the present method, this is a useful way of extracting data that is representative of the phenomenon for the purposes of the present chapter.

### 4.1.3 Oblique argument data

Using Python 3.6., verbs with a relative frequency over 1 per million words (pmw) were extracted, which co-occurred with nominals marked in one of the six spatial case suffixes in the same clause<sup>21</sup> in more than 50% of their instances. 200 such verbs were found.

In order to determine the suitability of these verbs, 50 random clauses per each verb were manually examined<sup>22</sup> (n = 10,000). During this process, 78 verbs were excluded from the dataset. Verbs were excluded for the following reasons:

---

<sup>21</sup> A clause is an automatically segmented part of a sentence that ideally has one finite verb form and the information pertaining to it (arguments and modifiers). In the present study it is the result of a clause splitting algorithm used in the Balanced Corpus of Estonian. Segmentation relies on punctuation, finite verb forms and various particles.

<sup>22</sup> This selection included only finite verb forms. This means that infinitive and converb verb forms were automatically excluded (*-da, -ma, -tud, -vat, -mas, -mast, -maks, -mata, -des*).

- If the final co-occurrence frequency dropped to below 50% due to clauses which were determined to not contain a complement upon manual checking. This could happen due to various reasons:
  - Participial forms were detected among clauses (*ratast parandanud laps* ‘a child that had fixed the bike’). This removed *kiiritama* ‘emit radiation’ and *taasavama* ‘reopen’ from the list.
  - A verb co-occurred with adpositions in more than half of its instances. For instance, *pookima* ‘graft onto’ frequently occurred with the postpositions *külge* and *juurde* ‘onto’ in addition to an allative argument.
- A verb was restricted to either a single or very few argument lexemes. For instance, *mõlkuma* ‘ponder’ is one of several verbs which only occurred with a small number of mental lexemes, e.g. *peas* ‘in head’, *mõttes* ‘in thought’ or *meeles* ‘in mind’. The same mental nominal tendency applied to *vasardama* ‘hammer’ and *turgatama* ‘crop up’. *Noteerima* ‘list’ was removed as it only occurred with *börsi-l* stock.market-ADE ‘on the stock market’.
- A verb was restricted to only one inflectional form. For instance, almost all instances of *orienteerima* ‘orient towards’ occur in participial/perfect forms (*kasvule orienteeritud* ‘oriented towards growth’). The same applies to *tingima* ‘condition’ (*sellest tingitud* ‘due to this’).

Finally, it was observed that some of the remaining 122 verbs in the oblique argument verb list frequently co-occurred with constituents, the role of which appeared to be physical locations, e.g. *patrullima* ‘patrol’:

- (45) Tänava-l patrulli-vad politseinik-ud  
 street-ADE patrol-PRS.3PL police.officer-NOM.PL  
 ‘Police officers are patrolling [on] the streets’

From the perspective of this study, it is desirable to exclude such verbs, because whether or not this constitutes argument structure is the topic of a whole other study and does not fit in the scope of the present study. We excluded verbs with frequent spatial adverbials by means of two variables. All 122 verbs were lexically coded for the semantic role of the oblique argument (constant per each verb). In addition, the 50 random clauses of each of the 122 verbs, where verbs co-occur with overt arguments, animacy of the argument referent was coded for each clause. Animacy was coded on four levels: human, other animate, inanimate concrete and inanimate abstract. We expected physical locations to pattern with the ‘inanimate concrete’ category. Animacy was useful as verbs frequently co-occurring with abstract locations were kept in the dataset.

Following this, a verb was excluded from the dataset if it filled both of the following two conditions: 1) its complement was coded as a Goal, Source or

Location, and 2) the animacy type of the referent of its argument was ‘inanimate concrete’ in more than half of the clauses including that verb. Altogether, 18 verbs were removed due to these conditions. These verbs include *lahustuma* ‘become dissolved in’, *nõjatuma* ‘lean on’, *sisenema* ‘enter’ and *patrullima* ‘patrol somewhere’.

In the rest of the dataset (n = 104 verbs), physical locations are infrequent. Only 10% (n = 12) of these verbs co-occur with a physical location somewhat frequently, i.e. in 30–50% of their instances (they are included in the final dataset of 104 verbs). For instance, *eralduma* (‘separate from’) remains in the oblique argument dataset, because approx. 55% of its complements are inanimate abstract Sources where the verb’s meaning makes no reference to actual locations. Instead, it is closer to ‘become different from’ (46):

- (46) kultuur eraldu-b kunsti-st  
 culture.NOM separate-PRS.3SG art-ELA  
 ‘Culture separates from art’/‘Culture becomes distinguished from art [in some context]’

In the rest of its instances (45%), however, *eralduma* co-occurs with physical locations, e.g. describing gas emanating from factories.

### The final oblique argument verb and clause datasets

The final oblique argument verb sample contains 104 verbs. The dataset is freely available (Aigro, 2022b). The final oblique argument clause sample includes 50 random clauses per verb, extracted from the Balanced Corpus of Estonian. For verbs that had less than 50 instances (but more than 15, as 1 pmw was set as an inclusion threshold), the total number of instances was used. All clauses in the dataset include an overt oblique argument (n = 4,502 clauses). Both the verb and clause sample are described in Table 5:

**Table 5:** Oblique argument verb and clause samples. Each verb occurs with a single spatial case in more than 50% of its instances. Mean clauses per verb = 43.3.

Case	Number of verbs in the dataset	Number of observations (% of dataset)
Allative (‘onto’)	39	1,664 (37.0%)
Adessive (‘on’)	5	250 (5.6%)
Illative (‘into’)	12	483 (10.7%)
Inessive (‘in’)	10	460 (10.2%)
Elicative (‘out of’)	38	1,645 (36.5%)
<b>Total</b>	<b>104</b>	<b>4,502</b>

No verb co-occurred with ablative arguments (‘from on top of’) frequently enough to be included in the dataset. We intentionally set no restrictions on the types of verbs, allowing the dataset to contain two-place verbs with subject-like Experiencers (47a) as well as two- and three-place verbs with object-like Themes (47b)–(47c).

- (47) a. Ta-l ei õnnestu-nud saa-da soovi-tud haridus-t  
 3SG-ADE NEG succeed-CNG.PST get-INF1 want-PPT education-PAR  
 ‘S/he did not succeed in obtaining the desired education’
- b. Ma ei hooli eriti sünnipäeva-de-st  
 1SG.NOM NEG care.CNG.PRS much birthday-PL-ELA  
 ‘I don’t care much about birthdays’
- c. Psühhiaater süüdistas stressi kasv-us elutempo  
 psychiatrist.NOM accuse-PST.3SG stress.GEN increase-INE life.pace.GEN  
 kiirenemis-t  
 acceleration-PAR  
 ‘The psychiatrist blamed the increase in stress on the accelerating pace of life’

### Rätsep’s valence list

Before turning to the described method of extracting verbs from the corpus, an attempt was made to obtain data based on an existing extensive valence list for Estonian verbs, compiled by Rätsep (1978). From his manuscript, all verbs and their valence frames were manually extracted, for which Rätsep proposed that their valence involves either an obligatory or a voluntary element marked in a spatial case, which is not a location. The list includes one-place, two-place, three-place and four-place verbs — altogether 1,116 valence frames occurring with 536 verbs. While Rätsep’s list also included verbs which were marked as frequently occurring with a spatial adverbial, they were not included.

It was finally determined that Rätsep’s list could not be used for the verb study. The main issue was that it appeared to exclude a number of key verbs, e.g. *suhtuma* ‘have an attitude towards’/‘relate to’. In addition, it contained a number of verbs which may be regarded as rare with the assigned case frame. For instance, *alarmeerima* ‘alarm’ is analysed as selecting an obligatory relative argument (*alarmeeris mind ohu-st* alarm-PST.3SG 1SG.PAR danger-ELA ‘S/he alarmed me about danger’).

Furthermore, when Rätsep’s 536 oblique argument verbs were mapped with the frequencies with which they co-occurred with spatial cases, only 65 verbs exceeded the 50% argument co-occurrence rate and the 1 pmw frequency requirement. Conversely, the corpus method rendered 200 such verbs. Hence, it was decided that the verb sample should be extracted empirically from the corpus. This also has the advantage of reflecting actual language use, rendering more reliable data for the purposes of this study.

#### 4.1.4 Canonical argument data

Canonical argument data includes verbs which select an object in one of the canonical object cases, i.e. partitive, genitive and nominative. It was compiled in a manner analogous to the oblique complement dataset.

In order to obtain a verb sample of similar size to that of the oblique set ( $n = 104$ ) and to render a dataset that is appropriately sized for manual inspection of verb instances, verbs with a relative frequency of 100 pmw or more were first extracted ( $n = 260$  verbs). These verbs were then mapped with their co-occurrence frequencies with nominals marked in partitive and genitive. Although the Estonian differential object marking paradigm also allows for nominative direct objects (see chapter 3.3.2), it is the least frequent object case and a common subject marker, which is why it was omitted as a condition from this verb extraction process. The final clause dataset may nevertheless include nominative objects.

We then selected verbs which co-occurred with partitive and genitive nominals most frequently, i.e. which always occur with either a partitive and/or a genitive constituents in the same clause ( $n = 222$  verbs). A sample of 50 random clauses per each of these verbs was compiled ( $n = 11,100$ ). Each clause was manually coded for the presence or absence of direct objects. This is because structural cases have a range of other functions than just marking objects (e.g. genitive marking possession or nominal selection by adpositions, partitive occurring in quantifier phrases, to name a few). Hence, the occurrence of a case token does not imply the occurrence of a direct object — its status would need to be confirmed separately. Furthermore, there are various syncretism patterns between nominative, genitive and partitive morphology, which affects the quality of their automatic annotation in the corpus.

A verb was included in the final verb sample if its object co-occurrence rate was equal to or above 50%. After the manual check, the final verb sample contained 128 verbs. The dataset is freely available (Aigro, 2022b).

The final clause sample contains clauses that include an overt object ( $n = 5,364$  clauses, from the total of 6,400 clauses or 50 clauses per verb). It is described in Table 6:

**Table 6:** Canonical object dataset, including 128 verbs (mean clauses per verb = 41.9). The dataset is freely available (Aigro, 2022b).

Case	Number of obs.	% of all direct objects
Partitive object	3,530 (55.2%)	65.8%
Genitive object	1,068 (16.7%)	19.9%
Nominative object	766 (11.9%)	14.3%
<b>Total</b>	<b>5,364</b>	<b>100%</b>

## 4.2 Study 1: The lexical transitivity of verbs

This section will focus on the first of the three research questions presented in the beginning of the present chapter: Are oblique argument structures assigned to lexically less transitive verbs than canonical argument structures? If yes, in what way are these verbs less transitive? Lexical transitivity will be explored in the two verb samples described in sections 4.1.3 and 4.1.4.

### 4.2.1 Method

Operationalising transitivity into a measurable variable is not a straightforward task. The transitivity parameters of Hopper and Thompson (1980) (also reflected in the Proto-role entailments of Dowty (1991)) have been helpful for discovering commonalities underlying non-canonical case usage in a cross-linguistic perspective. These parameters are said to affect the overall transitivity of a clause, with some pertaining to verbs (kinesis, punctuality, number of participants), some to subjects (agency), objects (affectedness and individuation), and others to the entire clause (negation and aspect). However, due to their overall number it is more straightforward to use these parameters to explain individual examples of differential case marking (Kittilä, 2008; Lestrade, 2010) than to quantitatively demonstrate which of them are stronger factors in the assignment of canonical and oblique argument structures to verbs.

In addition, these parameters exhibit a great deal of covariation (Malchukov and De Swart, 2008, 340). This, however, ought not to be viewed as a disadvantage for parameters as a system, but informative of what they really measure. A number of parameters proposed by (Hopper and Thompson, 1980) seem to pattern in the same way for stative verbs, including ‘kinesis’ (lack of movement), ‘punctuality’ (lack of natural temporal end points), ‘volitionality’ (lack of willfully acting subjects), ‘agency’ (lack of Agents), ‘affectedness of O’ (no real effect on objects), ‘individuation of O’ (non-individuated objects) and ‘aspect’ (atelicity). Hence, covariation seems to form a web with stativity in the middle — an underlying common factor which many parameters simultaneously indicate. This is also reflected in the analysis of Tsunoda (1981, 396), where stativity is viewed as underlying the entire transitivity-based verb type hierarchy, constituting the variable measured by other features. In addition, Onishi (2001) mentions stativity as one of the three main factors affecting the choice of case frame, along with lack of control (volitionality) and modality (irrealis vs realis).

The observation that stativity and transitivity are somewhat parallel features is useful for study searching for ways to operationalise transitivity in corpus studies. Similarly to transitivity, stativity, too, has been analysed as having compositional nature. However, differently from transitivity, which is a wide collection of measures, which may end up in fixed patterns anyway, stativity may be viewed as more compact, made up of fewer levels.

As a lexical property of verbs, lexical stativity broadly refers to two different notions: the non-dynamicity and lexical atelicity of states-of-affairs (events and states) (Vendler, 1957; Lakoff, 1966; Onishi, 2001). Non-dynamicity means that in entirely stative situations participants do not move. Atelicity means that these events/states-of-affairs do not have inherent temporal end-points. These two conditions would render verbs such as *tmõistma* ‘understand’, *kuuluma* ‘belong’, but also non-dynamic lexically atelic verbs such as *istuma* ‘sit’ and *ootama* ‘wait’.

However, in the classic Aktionsart system, differentiating between activities, states, accomplishments and achievements, the class of states is restricted in other ways as well. They are also identified by their lack of continuous forms (Vendler, 1957; Lakoff, 1966, 146) (\*‘I was just belonging to a group while something else happened.’). They are semantically defined as the grey area between properties and actions (Vendler, 1957, 152). This means what while the set includes states such as *sõltuma* ‘depend on’ and *kuuluma* ‘belong’, it excludes stative verbs depicting volitional events, e.g. *lamama* ‘lie’ and *istuma* ‘sit’.

Hence, certain types or degrees of stativity exist, which are not captured by the Aktionsart system of Vendler (1957), which views one such type (volitional stative verbs such as *istuma* ‘sit’) as part of the activity group, analogous to *jooksma* ‘run’. In a study focused on outlining the role of stativity in another phenomenon (argument structure assignment), one may wish to take a wider perspective in order to investigate not only the role of states, but also the role of individual aspects of stativity (lexical non-dynamicity and atelicity).

A more comprehensive system of stativity has been developed by Maienborn (2003). It outlines three categories for verbs expressing stativity: K-states, D-states and non-states. K-states (or kimian states<sup>23</sup>) refer to non-volitional states, expressing properties and relationships rather than events, such as *sarnanema* ‘resemble’ (mapping on the state category in Vendler, 1957). D-states (or davidsonian states<sup>24</sup>) refer to stative events, such as *istuma* ‘sit’ and *ootama* ‘wait’ (Maienborn, 2003, 2008, 2019; Rothmayr, 2009). Non-states refer to events which are either telic or include at least some degree of movement (*parandama* ‘fix’, *jooksma* ‘run’) (Fábregas and Marín, 2012).

The main formal distinction between K-states and D-states lies in the fact that D-states entail a hidden event argument, which was introduced by Davidson (1967). D-states therefore qualify as Davidsonian eventualities, which are “particular spatiotemporal entities with functionally integrated participants” (Maienborn, 2008, 109). K-states (*omama* ‘own’, *armastama* ‘love’, *näljane olema* ‘be hungry’), however, are “abstract objects for the exemplification of a property P at a holder x and a time t” (Maienborn, 2008, 113). This means that instead of expressing events including acting participants, K-states express properties of par-

---

<sup>23</sup> named after Kim (1969)

<sup>24</sup> following Davidson (1967)

ticipants. This is based on the observation that K-states fail the following characteristics of eventualities (Maienborn, 2008):

1. Eventualities are perceptible ('I saw her sit there', \*'I saw her resemble a horse')
2. Eventualities are located in space and time, while K-states are only located in time ('Clothes are drying on the balcony', \*'I am hungry in the kitchen')
3. Eventualities combine with manner adverbials, instruments, comitatives, etc. ('She is sitting there with her sister', \*'She is hungry with her sister')

As mentioned, in terms of Aktionsart, K-states broadly align with the classic notion of states which lack continuous forms (Vendler, 1957; Lakoff, 1966). D-states, on the other hand, would belong under activities.

All in all, the benefit of using the framework of Maienborn (2008) to assess the transitivity distinction between canonically and non-canonically marking verbs is multi-fold. First, its inclusive nature and internal gradience allows us to map the transitivity distinctions in more detail, as it links the class of properties (K-states) to the rest of the verb space through the grey area of stative events (D-states). Second, due to its fewer levels, the three-level stativity system of Maienborn (2008) renders a more coherent and intuitive gradient in transitivity than the wide range of transitivity factors (Hopper and Thompson, 1980).

Third, Maienborn's system is especially useful for an argument structure study, because it explicitly and prominently encodes lexical aspect, which has been proposed as the main mediator between semantics and syntax in some approaches (Tenny, 1992; Verkuyl, 1999; Kratzer, 2004; Tamm, 2004). For instance, Tenny 1992 uses aspect as the very basis of internal argument definition, suggesting that direct objects constitute entities that measure out the event through change. They can also delimit events by encoding their temporal end points ('destroy'), although they need not do that ('push'). However, Tenny only investigates the role of aspect in canonical argument structure in English. Hence, a large group of verbs which are expected to be semantically rather distinct from canonically marking verbs has been left out of the analysis. If aspect really plays fundamental role in argument realisation, one would expect canonical and non-canonical complement verbs to have distinct lexical aspect patterns, given their distinct realisation patterns.

Finally, unlike the transitivity parameters in (Hopper and Thompson, 1980), which pertain to various parts of clauses, Maienborn's system is specifically designed to assess the lexical semantics of verbs. Hence, it is more suitable for a study focusing on verbs selecting oblique non-canonical argument structures rather than clause-dependent case alternations. Hence, the two verb samples will be compared in terms of their proportions of K-states, D-states and non-states.

These three state types were coded based on four variables: acceptability as a complement of the perception verb *nāgema* 'see', verb type, lexical dynamicity and lexical telicity. The following sections will discuss the coding of each of the four variables.

## Acceptability judgment task

Maienborn proposed that K-states, which constitute properties rather than events, cannot be observed, meaning that verbs expressing K-states tend not to occur as perception verb complements, e.g. \*‘I saw tomatoes weigh 1 kg’ (Maienborn, 2008, 110). In order to test which verbs in our two datasets resist this phenomenon, an acceptability judgment task was organised.

In this experiment, canonical and oblique argument verbs were presented in a sentence context where they occur as complements of the perception verb *nägema* ‘see’.<sup>25</sup>

- (48) Nägin Andrus-t voodi-l muusika-t kuula-mas  
see.PST.1SG Andrus-PAR bed-ADE music-PAR listen-CONT  
‘I saw Andrus listening to music on the bed’

The experiment included 232 experimental items (sentences) — 104 of which contain oblique verbs and 128 canonical object verbs. The 232 stimuli were divided into two lists. 49 fillers were added to list 1 and 48 fillers were added to list 2. Fillers were designed to be distinct from stimulus sentences in order to mix the overall set of sentences. They all started with subject constituents, followed by finite verb forms.

Each experimental item was in the same form (48). It contained: 1) the perception verb (*Nägi-n* see.PST-1SG ‘I saw’), 2) target verb in continuous form (*kuula-mas* listen-CONT ‘listening to’), 3) subject of the target verb and the object of the finite verb (*Andrus*) and 4) a modifier (e.g. *voodi-l* bed-ADE ‘on the bed’). Canonical object verbs and some oblique argument verbs also include 5) an argument of the target verb (*muusika-t* music-PAR ‘music’ in (48)). Visual modality (the verb *nägema* ‘see’) was chosen to represent perception verbs, because it is the primary sense modality in Estonian, given that vision verbs are more frequent than other perception verbs (Proos, 2021).

The other four lexemes (subjects, complements and adjuncts) were chosen based on the highest collocation log-likelihood data.<sup>26</sup> For instance ‘music’ was chosen for ‘listen’ in (48). If all main collocators of a verb were personal pronouns, a proper noun in the form of human name was chosen instead as subject.

26 participants completed the offline task for list 1 and 17 participants for list 2. Participants rated each sentence on a 7-point Likert scale, based on how acceptable the sentence appeared to them (1 = not acceptable, 7 = highly acceptable).

---

<sup>25</sup> It has long been highlighted as one of the main distinctions between states and non-states that the former are not accepted with continuous tense inflection, e.g. \*‘I am knowing.’ (Vendler, 1957, 148). Maienborn’s K-states broadly align with Vendler’s (1957) states. The stimulus sentence in (48) includes the Estonian verb in continuous form. Hence, testing sentences such as (48) somewhat accidentally identifies K-states both semantically and structurally.

<sup>26</sup> Data originates from the database of collocation information from the Estonian National Corpus: <https://korpused.keeleressursid.ee/clc/?mod=search>.

## Verb type

K-states were also identified based on verb types, which have been shown to apply to them in literature. It was coded on four levels: 1) experiencer verbs (*hoolima* ‘care about’, this group also includes various possessor verbs), 2) measure verbs (*koosnema* ‘consist of’) and 3) property verbs (*võlgnema* ‘owe’) and 4) other (*virutama* ‘hit’).

Experiencer/possessor and measure verbs were chosen as K-state indicators based on (Rothmayr, 2009), where these verb types are shown to systematically behave as K-states. Property verbs are added to this group based on the core characteristic of K-states in (Maienborn, 2008): “they are abstract objects for the exemplification of a property P at a holder x at a time t” (Maienborn, 2008, 113).

## Dynamicity and telicity

In order to code D-states and non-states, all verbs not classified as K-states were coded for lexical dynamicity (dynamic vs non-dynamic) and telicity (telic, atelic). A verb was deemed dynamic if it had at least one sense in which one or all participants need to physically move, even if only slightly (*lähenema* ‘approach’ is dynamic while *hoolima* ‘care about’ is not). Interaction verbs (*informeerima* ‘inform about’) were coded as dynamic to leave the label ‘non-dynamic’ strictly for zero-movement verbs. A verb was deemed telic if it entailed an inherent temporal end point in at least one of its senses. Hence, *sõltuma* ‘depend on’ is lexically atelic while *loobuma* ‘give up’ is telic.

## Coding states

In order to be coded as a K-state, a verb had to fulfil four conditions<sup>27</sup>:

1. It is not highly accepted as a perception verb complement (see first section in 4.2.1), i.e. its mean acceptability score on a 7 point scale was 4 or less
2. It is an experiencer/possessor verb, measure verb or a property verb
3. It is lexically non-dynamic
4. It is lexically atelic

D-states and non-states were coded among the verbs not classified as K-states. Following the definition that D-states constitute spatiotemporal eventualities which

---

<sup>27</sup> Another prominent K-state test is that K-states are expected to resist deverbal nominalisation (\*‘The resembling’) (Maienborn, 2008; Rothmayr, 2009). We examined the nominalisations formed with the suffix *-mine*, the most productive deverbal nominalisation suffix in Estonian (Kasik, 2015; Pilvik, 2019). Unexpectedly, however, oblique argument verbs actually showed higher nominalisation frequencies than canonical complement verbs in Estonian, when these counts were viewed in relation to the overall frequencies of verb lemmas. This casts doubt on nominalisation resistance as a universal K-state property. Instead, it appears to depend on the language-specific productivity of various derivational deverbal morphemes.

entail neither motion nor inherent temporal end points (Rothmayr, 2009), a verb was deemed a D-state if it was simultaneously coded as non-dynamic and atelic (e.g. *toetuma* ‘lean against’/‘be supported by’). The verbs which were neither K-states nor D-states were tagged as non-states.

### Issues with encoding

Encoding semantic variables, such as the three variables discussed above (dynamicity, telicity and verb type), inevitably introduces a degree of subjectivity into the study. Broadly, such variables are problematic in three ways.

First, binary semantic variables only reflect graded phenomena in a simplistic way. Hence, a binary notion of dynamicity naturally ignores the fact that motion, for instance, varies in intensity and duration. Similarly, a binary assessment of lexical aspect simplifies the wide variance in the temporal organisation of events.

Second, some verbs are more polysemous than others, meaning that their multiple senses can render their lexical forms ambiguous in various dimensions. Ambiguity can arise from the animacy of subjects (Rothmayr, 2009) — *järgnema* ‘follow’ is non-dynamic with an inanimate subject (*Esmaspäeva-le järgne-b teisipäev* monday-ALL.SG follow-PRS.3SG tuesday.NOM ‘Tuesday follows Monday’), while an animate subject gives it a dynamic reading (*Poiss järgne-b koera-le* boy.NOM follow-PRS.3SG dog-ALL ‘The boy follows the dog’).

Ambiguity can also arise from the semantics of the complement. For instance, *lähenema* ‘approach’ is dynamic with a concrete complement (*lähene-b rannale* approach-PRS.3SG beach-ALL ‘approaches the beach’). However, it is non-dynamic with an abstract complement (*lähene-b probleemil-e* approach-PRS.3SG problem-ALL ‘approaches the beach’).

Furthermore, ambiguity can depend on discourse contexts. For instance, *meenutama* ‘remind of’ can refer to an atelic property of things (*meenuta-d mu venda* remind-PRS.3SG my brother.PAR ‘You remind me of my brother’), but it has a more telic sense when it is used as an interaction verb (*meenuta-s mul-le meie vestlust* remind-PST.3SG 1SG-ALL 1PL.GEN conversation.PAR ‘S/he reminded me of our conversation.’).

Third, a number of verbs are highly abstract, with their dynamicity and aspect interpretations relying almost entirely on the overall, more complex meaning of the expression. This includes verbs, the objects of which include deverbal nominalisations. For instance, whether or not *jätkama* ‘continue’ is dynamic, depends on the type of verb it selects as its infinitival complement. *jätka-b jooksmis-t* continue-PRS.3SG running-PAR ‘S/he continues running’ is dynamic while *jätka-b istumis-t* continue-PRS.3SG sitting-PAR ‘S/he continues sitting’ is not.

These issues were addressed by opting for a conservative approach. A verb was coded as dynamic even if it entailed very little movement (e.g. the presupposed movement of one’s mouth in interaction verbs, *arutama* ‘discuss’). Similarly, a single telic sense led to a verb being tagged as telic. Finally, abstract

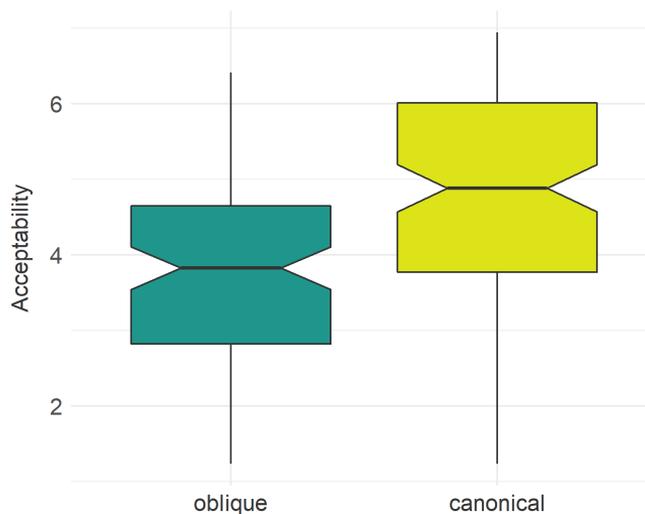
verbs were coded as dynamic, because, as mentioned, with some infinitival complements they render dynamic interpretations.

## 4.2.2 Results

This section will present the overall proportions of various types of states in the canonical and oblique argument verb samples. The following sections will present the results for each of the four variables coded to summarise stativity.

### Acceptability judgment task

An acceptability judgment task was organised (see section 4.2.1) to examine the acceptability of target verbs when they occur as complements of the perception verb *nägema* ('see'). K-states are said to differ from D-states in that they describing temporal but not spatial properties, meaning one cannot perceive them (Maienborn, 2008; Rothmayr, 2009). Hence, oblique verbs, which are hypothesised to be more stative and therefore include a higher proportion of K-states, are also expected to be less acceptable in such a context.



**Figure 1:** Acceptability ratings per verb group when a verb occurs as a complement of the perception verb *nägema* 'see' on a 7 point Likert scale;  $n(\text{oblique}) = 104$ ,  $n(\text{canonical}) = 128$

First, as seen in Fig. 1, oblique argument verbs are significantly less acceptable as perception verb complements than canonical argument verbs (oblique: mean = 3.9,  $sd = 1.3$ ; canonical: mean = 4.8,  $sd = 1.3$ ,  $W = 9158.5$ ,  $p < 0.001$ ). Second, oblique complement verbs are generally not acceptable in this structure, with a

mean judgment below 4 for the entire group. There are almost twice as many verbs with a mean rating of 4 in the oblique argument dataset than in the canonical argument dataset (56% vs 30%).

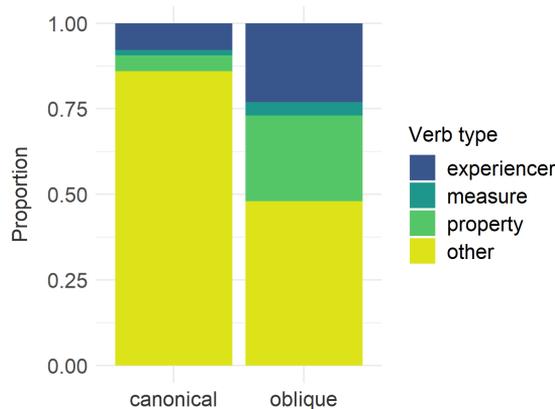
Illative argument structure stands out in the oblique set as it is the only structure, the verbs of which are generally acceptable as complements of a perception verb (mean = 4.4, sd = 1.33). Easily perceived illative verbs include *sukelduma* ‘dive in’/‘become immersed in’, *sekkuma* ‘involve oneself in’, *investeerima* ‘invest in’ and *mattuma* ‘become buried in’, as in (49):

- (49) Näg-in kuuri lumme mattu-mas  
 see-PST.1SG shed.PAR snow.ILL be.buried-CONT  
 ‘I saw the shed being buried in snow’

Canonical verbs are not statistically distinct from inessive and illative verbs, but they do significantly differ from relative verbs ( $p = 0.008$ ), allative verbs ( $p = 0.002$ ) and adessive verbs ( $p = 0.005$ ).

### Verb type

Verb types were coded for four levels: experiencer/possessor verbs, measure verbs, property verbs and the rest. Fig. 2 outlines their proportions in the two verb samples:

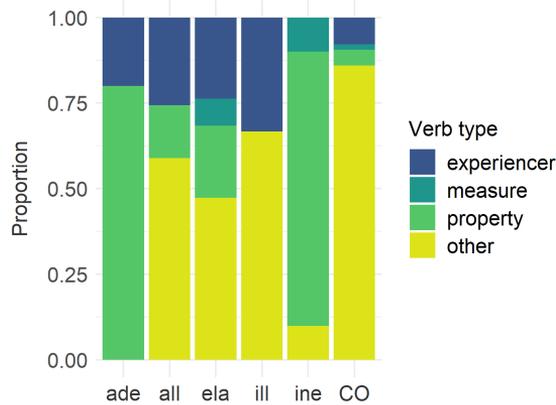


**Figure 2:** Verb type per dataset.

Of these four groups, only measure verbs appeared not to be sensitive to argument structure, being relatively rare in both samples. Both experiencer and property verbs were significantly more numerous in the oblique argument dataset (experiencer verbs in oblique vs canonical dataset:  $\chi^2 = 9.5037$ ,  $df = 1$ ,  $p = 0.002$ ; property verbs in oblique vs canonical dataset:  $\chi^2 = 18.239$ ,  $df = 1$ ,  $p < 0.0001$ ).

All in all, the types of verbs describe 52% of the oblique complement dataset and 15% of the canonical complement dataset.

Fig. 3 outlines verb type proportions in different argument structures:

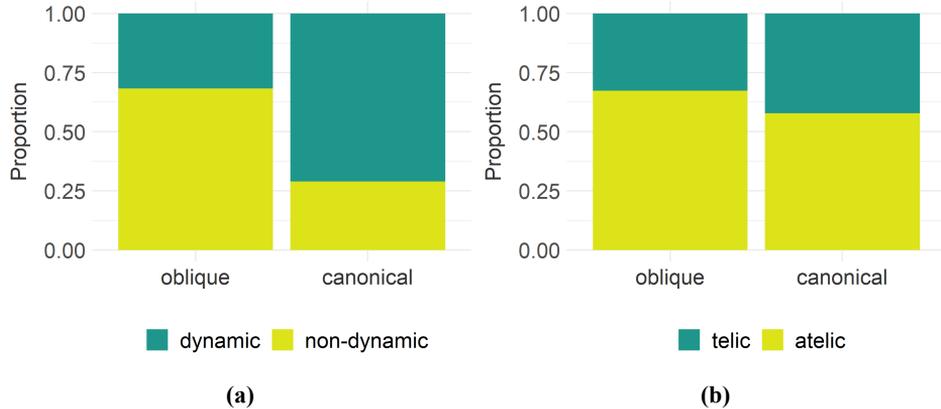


**Figure 3:** Verb type per argument structure (CO = canonical object in partitive, genitive or nominative).

Interestingly, in addition to the Experiencer verbs selecting allative and adessive arguments, similar proportion of such verbs are found among those governing elative (*hoolima* ‘care about’, *loobuma* ‘give up’, *tüdinema* ‘become tired of’) and illative (*armuma* ‘fall in love with’, *kiinduma* ‘become fond of’, *suhtuma* ‘relate to’/‘have an attitude towards’) (Fig. 3). Inessive and adessive are mostly selected by property verbs (*baseeruma* ‘be based on’, *avalduma* ‘be reflected in’), while illative stands out in terms of the largest proportion of ‘other’ verbs, as no illative verbs constitute measure or property verbs. Elative has the most diverse verb profile in the oblique argument dataset, similarly to transitive verbs.

## Dynamicity and lexical aspect

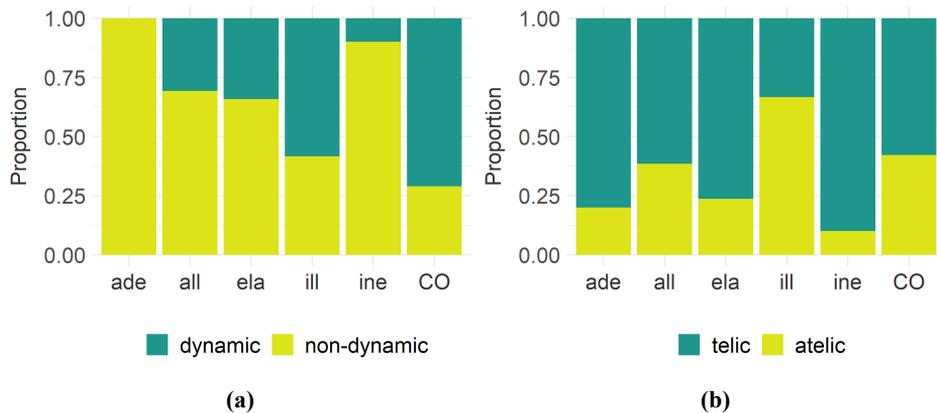
Fig. 4 summarises the proportions of dynamicity and aspect as lexical properties of verbs in both verb samples:



**Figure 4:** Lexical dynamicity (a) and aspect (b) in canonical and oblique verb samples.

All in all, oblique argument verbs are more distinct in terms of dynamicity, i.e. motion on the part of the participants of the states-of-affairs they describe, than in terms of lexical aspect, i.e. whether or not these states-of-affairs have natural temporal end points or not. In fact, there is no significant distinction between the telicity ratings of oblique and canonical argument verbs, but oblique verbs are significantly less dynamic than canonical argument verbs ( $\chi^2 = 34.168$ ,  $df = 1$ ,  $p < 0.0001$ ).

This means that there is a certain degree of asymmetry in the expression of stativity in oblique argument verbs. Stativity is traditionally viewed as made up of both lack of movement and lack of internal temporal boundaries. We used stativity to decompose transitivity in a more nuanced way. Results show that while oblique argument verbs entailing less motion, these verbs are not less transitive in terms of temporal organisation, as measured by lexical telicity.

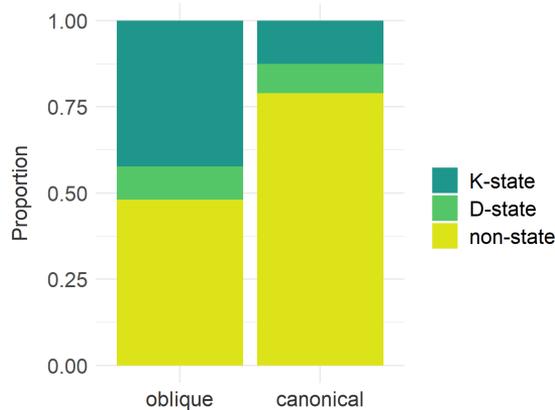


**Figure 5:** Lexical dynamicity (a) and aspect (b) per argument structure (CO = canonical object in partitive, genitive or nominative).

Fig. 5 shows a number of interesting structure-based differences. First, verbs in the illative structure stand out in terms of being lexically most dynamic and telic among the verbs in oblique structures. The dynamic, telic illative-selecting verbs of the type ‘other’ in the previous section were likely the reason for their above average acceptability ratings as well. Most illative verbs had at least one dynamic sense, including the verbs with the highest acceptability ratings, e.g. *sööbima* ‘become etched in’ and *suubuma* ‘end up in’. Most illative verbs are lexically atelic, this being the only oblique argument verb group which is on average more telic than canonical argument verbs.

## Lexical-semantic stativity

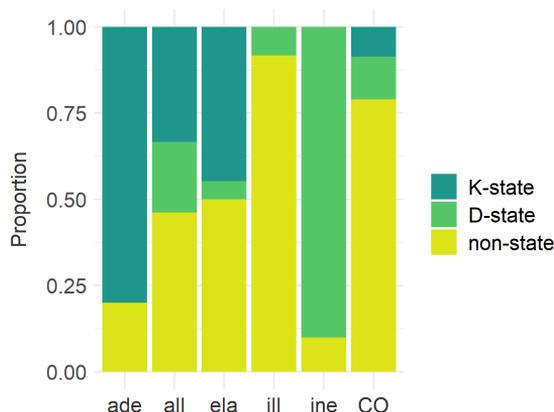
While the four variables described in section 4.2.1 are informative of the distinctions between various argument structures, their main function is to provide the basis for coding K-states, D-states and non-states in both verb samples. They are summarised in Fig. 6:



**Figure 6:** Verb state status per dataset.

Following the criteria outlined in section 4.2.1, 44 verbs were coded as K-states in the oblique argument dataset (42%). The same conditions rendered significantly fewer K-states in the canonical verb group ( $n = 16$ , 12%,  $\chi^2 = 25.057$ ,  $df = 1$ ,  $p < 0.0001$ ). D-states, i.e. non-dynamic atelic verbs constitute 22.1% of the oblique argument dataset and 11.7% of the canonical argument dataset, which is not a statistically significant distinction.

Finally, oblique argument verbs that were coded as non-states ( $n = 50$ ) include a significantly higher proportion of non-dynamic verbs (34%) than the non-state verbs selecting canonical objects ( $n = 101$ , 10%,  $\chi^2 = 11.637$ ,  $df = 1$ ,  $p = 0.0006$ ). These two verb groups do not differ in terms of telicity.



**Figure 7:** Verb state status per governed case. (CO = canonical object in partitive, genitive or nominative).

Unsurprisingly, illative-selecting verbs which were shown to differ from the rest of the oblique argument verbs in terms of all four variables, also differ in terms of state type proportions. The single stative verb coded in the illative group is the D-state *suhtuma* ‘relate to’/‘have an attitude towards’. All other illative-selecting verbs are coded as non-states, meaning they are either dynamic and/or telic.

Verbs occurring in adessive and inessive structures are also interesting, because the high property verb proportions demonstrated in Fig. 3 have translated into high K-state proportions in Fig. 7. For both verb groups, this is due to a number of verbs which constitute near-synonyms. The adessive-selecting synonyms make up almost the entire adessive dataset, accounting for 4 verbs out of 5 (excluding *õnnestuma* ‘succeed in’):

- *baseeruma* ‘be based on’
- *rajanema* ‘be founded on’
- *lasuma* ‘be based on’/‘lie on’
- *põhinema* ‘be based on’

The study found 10 verbs that select inessive arguments. Six of them may be considered to be almost synonymous:<sup>28</sup>

- *avalduma* ‘be expressed in’

<sup>28</sup> In addition, another three inessive-selecting verbs are semantically very close to them: *peituma* ‘hide in’/‘be reflected in’, *sisalduma* ‘be contained in’ and *pesitsema* ‘nest in’/‘be contained in’. The only inessive verb not belonging in the same semantic space as the others is *süüdistama* ‘accuse in’.

- *figureerima* ‘figure in’
- *kajastuma* ‘be expressed in’
- *peegelduma* ‘be reflected in’
- *seisnema* ‘constitute’/‘be reflected in’
- *väljenduma* ‘be expressed in’

This indicates that adessive and inessive are both used in argument structures that express a stative relationship (or a property of something) where indications of something can be observed in another context (*seadus baseeru-b eetika-l* law.NOM be.based-PRS.3SG ethics-ADE ‘the law is based on ethics’) These near-synonymous verb groups will be discussed in the context of syntactic productivity in section 5.4.

### Pragmatic stativity

While oblique argument verbs are more stative than canonical object verbs in terms of lexical-semantic stativity (section 4.2.2), the canonical dataset appears to include verbs that express stativity pragmatically.

Section 4.2.1 discussed subject animacy as one of the factors which may lead to ambiguity in verbs. It showed that a verb may have a stative interpretation with an inanimate subject referent and an active interpretation with an animate subject (‘The wall surrounds the town’/‘The army surrounds the town’). The canonical object verb sample includes 20 verbs which become stative with inanimate subjects. They are outlined below:

- *ühendama* ‘connect’
- *katma* ‘cover’
- *täitma* ‘fill’
- *väljendama* ‘express’
- *tingima* ‘drive a bargain’/‘cause’
- *moodustama* ‘create’/‘constitute’
- *keelama* ‘forbid’
- *iseloostama* ‘characterise’
- *kirjeldama* ‘describe’
- *näitama* ‘show’
- *puudutama* ‘touch’
- *suunama* ‘point towards’
- *rõhutama* ‘emphasise’

- *iseloostama* ‘characterise’
- *meenutama* ‘emphasise’
- *kaitsema* ‘protect’
- *kehtestama* ‘establish’
- *piirama* ‘surround’
- *tõestama* ‘prove’
- *sundima* ‘force’/‘cause’

These verbs have K-state interpretations with inanimate subjects (50a) and non-state interpretations with animate subjects (50b):

- (50) a. Müür kaitse-b linna  
 wall.NOM defend-PRS.3SG TOWN.PAR  
 ‘The wall defends the town’
- b. Sõjavägi kaitse-b linna  
 army.NOM defend-PRS.3SG TOWN.PAR  
 ‘The army defends the town’

Another three canonical object verbs are ambiguous due to polysemy where the active and stative senses are even more distinct and do not depend on the features of event participants:

- *maksma* ‘pay’/‘cost’
- *tähistama* ‘mark’/‘celebrate’
- *hindama* ‘assess’/‘appreciate’

Finally, active verbs in the canonical object dataset may also become stative with abstract arguments. This effect occurs whenever metaphorical meanings are introduced. For instance, *kandma* ‘carry’ is stative with an abstract object (*kanna-b vimma* carry-PRS.3SG grudge.PAR ‘carries a grudge’) and active with a concrete object (*kanna-b kotte* carry-PRS.3SG bag.PAR.PL ‘carries bags’). Virtually every atelic verb can be used with abstract objects, which then inadvertently introduce stativity into the clause.

In summary, while oblique argument verbs are lexically less transitive than canonical object verbs, this does not mean that the latter are used in a highly transitive way. A number of canonical object verbs were shown to alternate between stative and active interpretations. Many, if not all, become stative (and therefore less transitive) in an abstract context.

### 4.3 Study 2: Bleached use of case

This study aims to answer RQ2: To what degree can spatial semantics be detected in Estonian verbs occurring in oblique argument structures and how does this compare to canonical object verbs? Semantic roles were chosen as the way to indicate the presence or lack of spatial semantics in oblique argument verbs, as a lexical property of verbs. Section 4.3.1 outlines the method of coding roles while section 4.3.2 presents the results on the proportions of spatial semantics in Estonian oblique argument verbs.

#### 4.3.1 Method

Verbs in both samples were coded for the semantic role of their oblique argument or canonical object. The role of oblique arguments and canonical objects was coded as a lexical property of verbs. For instance, *lähenema* ‘approach’ always encodes a Goal argument, regardless of whether this is an abstract or a concrete Goal.

Because oblique argument verbs are less transitive (as was shown in section 4.2), coding semantic roles is not straightforward. This is because roles were originally designed with an implication of hierarchical structure, broadly referring to entities (Actor, Patient, Theme, Instrument) and locations for these entities (Goal, Source, Location, Path) (Fillmore, 1968; Kılıçaslan and Tuna, 2015). This system is therefore not easily applicable to verbs describing abstract properties or relationships between often inanimate entities (*sõltuma* ‘depend on’). Furthermore, in the dataset of the present study, spatial meaning, if present in the lexical semantics of verbs, is often abstract.

In order to accommodate various abstract types of relations, the study views roles as having rather broad definitions. A number of studies regard roles as concepts with prototypical structure, rather than semantically specific concepts (Hopper and Thompson, 1980; Dowty, 1991; Primus, 1999; Rissman and Majid, 2019). This means that each role is interpreted as a relatively wide notion. In addition, to keep the coding process simple and more easily replicable, only six roles were used — Experiencer, Theme, Recipient, Goal, Source and Location. Roles were independently coded by the author and two supervisors. Inter-coder agreement was 97.1%. Differences were solved through discussion. The following principles were followed in coding roles:

- A verb was coded as selecting a Goal, Source or Location if it has a sense in which it marks the concrete movement or location of something. Mostly, this describes verbs which alternate between concrete and abstract arguments, e.g. the elative argument of *eralduma* ‘separate from’ was coded as a Source and the adessive argument of *lasuma* ‘lie upon’ was coded as a Location. These roles were coded even if the spatial meaning is uncommon — e.g. *investeerima* ‘invest in’ was coded as selecting a Goal, even though it is uncommon for the investment to physically move to the target institution. As *baseeruma* ‘be based on’ is mainly used to mark abstract relations, it was coded as selecting a Theme rather than a Location.
- Experiencers were coded only for verbs which have a sense that expresses a mental activity of an animate entity. This included verbs such as *õnnestuma* ‘succeed in’ and *imponeerima* ‘like’.
- Recipients were coded instead of Goals if the verb encoded the role as predominantly human and made reference to a concrete transferred entity. This means that in at least one sense the verb must express an exchange between human participants. This includes *kinkima* ‘gift to’ and not *investeerima* ‘invest in’. In addition, *ulatama* ‘hand over to’ was coded as selecting a Recipient but *omistama* ‘assign to’ was not, because it cannot be used to describe transferring a concrete object (*doktorandile omistati teaduskraad* ‘The graduate student was assigned a degree’). Instead, it was coded as selecting an allative Goal. For the same reason, *andestama* ‘forgive’ was coded as taking an allative Theme.
- Patients were coded for verbs with at least one sense in which it expresses an event where the participant is physically changed. This includes *vähendama* ‘decrease’, *tapma* ‘kill’ and *jooma* ‘drink’. Because mere affectedness was not regarded as sufficient, *virutama* ‘hit’ and *lööma* ‘hit’/‘kick’ were coded as taking Themes.
- Theme is a role that was used to describe a participant of state-of-affairs that is not physically changed by the event, bears no spatial relation to the event. In this study, Themes also includes what sometimes are labelled Stimuli. For instance, both *ostma* ‘buy’ and *armastama* ‘love’ were coded as selecting Theme arguments.

### 4.3.2 Results

In general, the two types of verbs are distinct in terms of the number of different roles they mark, as well as the degree of semantically bleached roles. Canonical object verbs mark Themes and Patients in a higher proportion than oblique argument verbs, the difference being mostly made up by spatial roles in the oblique dataset (see Fig. 8).

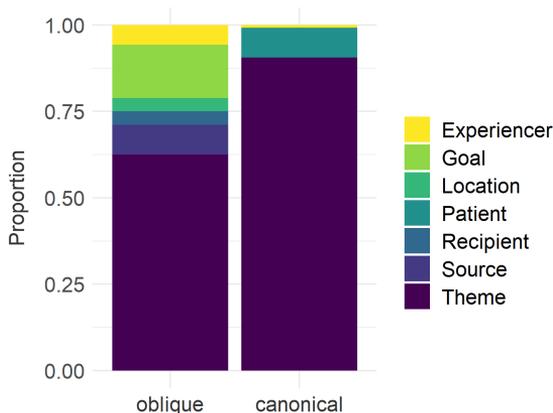


Figure 8: Semantic role per dataset.

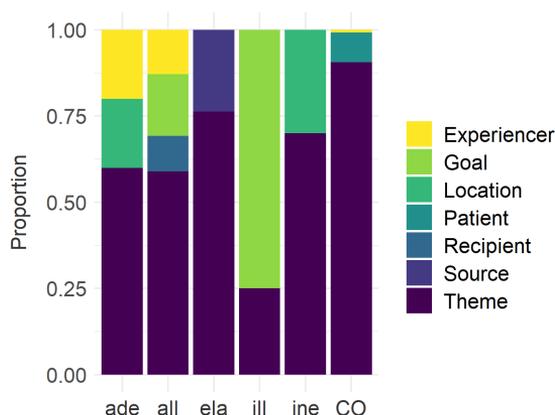
However, it is a remarkable finding that the two datasets are similar in terms of mostly being made up of verbs which mark semantically entirely bleached roles. Most verbs in both groups make no semantic reference to locations or very particular types of participants, such as Experiencers or Recipients, but they mark participants of events/states coded as Themes.

Hence, in this sample, only a minority of oblique argument verbs entail spatial semantics ( $n = 29$ , 27.8%). These verbs include illative Goals (*inkorporeer-ima* ‘incorporate into’), inessive Locations (*peituma* ‘hide in’) and allative Goals (*osutama* ‘point at’). About two thirds of oblique verbs ( $n = 65$ , 62.5%), however, encode Themes (e.g. the allative Theme of *alluma* ‘obey’ (*allu oma ema-le* obey.IMP.SG REFL mother-ALL ‘obey your mother’).

These results also address several assumptions made about canonical object verbs. For instance, following transitivity parameters (Hopper and Thompson, 1980), prototypical objects are assumed to be physically affected by the event described by the verb, i.e. constitute Patients. Among the 128 verbs in the canonical dataset, however, only a minority select such objects ( $n = 11$ , 8.6%, e.g. *suurendama* ‘increase’). Instead, similarly to the oblique argument dataset the canonical argument dataset is largely made up of verbs selecting Themes (e.g. *uurima* ‘investigate’), which are not physically affected by events.

## Semantic roles per argument structure

Fig. 9 outlines the semantic roles expressed in each argument structure:



**Figure 9:** Semantic role per governed case (CO = canonical object in partitive, genitive or nominative).

Four out of five spatial cases encode spatial meaning in less than half of the verbs which select them. Elative marks Themes the most frequently. Sources are only encoded by 23.7% of verbs selecting elative arguments (e.g. *järelduma* ‘be concluded from’). Allative marks Goals to an even lesser degree (17.9%, *lähenema* ‘approach’), the rest including Experiencers and Recipients in addition to Themes. This means that in terms of the degree of structural function represented by the present datasets, allative marks the widest variety of semantic roles, followed by adessive, which also marks Experiencers in addition to Themes and spatial roles.

In addition, verbs with allative Themes (n = 23) are interesting, because they have received less attention than allative Recipients and Experiencers. Verbs selecting allative Themes make up almost two thirds of the allative dataset in our study. They include *eelnema* ‘precede’, *vihjama* ‘hint at’, *võlgnema* ‘owe’ and *alluma* ‘obey/succumb’.

Although verbs selecting inessive and adessive mostly encode Themes in our dataset, this cannot be straightforwardly regarded as an indication of their semantic generality and multifunctionality. As was shown in section 4.2.2, most verbs in both datasets constitute near-synonyms. Semantic roles do show slight nuances in terms of how they differ in locative semantics. For instance, only one of the four synonymous adessive-governing verbs was deemed to entail spatial semantics (*lasuma* ‘lie on’/‘be based on’) as it was the only one with an available concrete location sense. The same applies to three of the nine near-synonymous inessive-governing verbs (*peituma* ‘hide in’/‘be reflected in’, *pesitsema* ‘nest in’/‘be contained in’, *sisalduma* ‘be contained in’). Hence, even though

several of the verbs selecting inessive and adessive arguments code these cases as bleached from spatial semantics, these argument structures are nevertheless used with verbs that form a semantically coherent group.

Finally, illative is the only case which occurs in structural position with non-bleached semantics. Only three illative verbs were coded as selecting Themes: *ar-muma* ‘fall in love with’ *kiinduma* ‘become fond of’ and *suhtuma* ‘relate to’/‘have an attitude towards’.

In summary, most verbs in our dataset which select spatial cases do not generally mark spatial semantics — not even in only one or few of their several senses.

#### 4.4 Study 3: Usage patterns of argument structures

This section will focus on RQ3: Does the usage of verbs with oblique argument structures differ from that of canonical argument structures? If yes, in what way? It is an explorative study where usage patterns are investigated by means of the main syntactic, morphological and semantic variables describing clauses. The two clause samples are described in sections 4.1.3 and 4.1.4.

Section 4.4.1 will describe the methodology and variables used in the subsequent analysis. Section 4.4.2 describes the output of mixed logistic regression analysis, outlining the main distinctions in the ways the two verb groups are used. Finally, section 4.4.3 investigates the usage patterns of individual argument structures.

##### 4.4.1 Method

We fit a mixed logistic regression model to the data, using the lme4 package for R (Bates and Maechler, 2009).

The dataset is made up of clauses including oblique and canonical argument verbs (described in section 4.1). It only includes clauses with an overt argument and it is balanced for verbs. For oblique argument verbs, it including 50 random instances per verb, or less for verbs lower in overall corpus frequency. For canonical object verbs, clauses with overt objects were extracted from a sample of 50 instances per verb. Table 7 presents the clause data previously shown in Tables 5 and 6:

**Table 7:** Clause dataset.

Verb type	Verbs	Mean clauses per verb	Overall clauses
Oblique argument verbs	104	43.3 (sd = 11.9)	4,502
Canonical argument verbs	128	41.9 (sd = 7.0)	5,364
Total	232		9,866

The dependent variable of the model is *type of structure* (‘oblique’ vs ‘canonical’). Fixed effects include 11 independent variables, described in Table 8. On

the one hand, these variables include factors that pertain to argument information in a clause. For instance, Estonian uses word order in addition to morphosyntax to convey argument information (Lindström, 2005). In addition, subjects and objects are generally expected to show animacy distinctions (de Hoop and de Swart, 2009; Lestrade, 2010), which could give rise to pronominalisation and number distinctions. Lexeme-based NP properties might describe a distinction between arguments and adjuncts. For instance, phrases referring to arguments are expected to be shorter and case-marked in a higher proportion than adjuncts (given, for instance, their lack of adpositions in the sample used in this study). One might also expect to find them occurring closer to the verb than adjuncts.

In addition, as clauses are essentially extracted based on verbs, main verbal inflections were included as factors (tense and negation), as well as register to outline verb functionality and usage.

**Table 8:** Variables in mixed logistic regression. ‘Complement’ signifies argument phrases, both oblique and canonical.

Variable	Type	Levels
type of structure	categorical	oblique, canonical
order	categorical	complement precedes the verb, complement follows the verb
pronoun	categorical	complement is a pronoun, complement is not a pronoun
number	categorical	complement is singular, complement is plural
tense	categorical	finite verb is in present, past, perfect or pluperfect tense
negation	categorical	finite verb is affirmative, finite verb is negative
register	categorical	clause originates from journalism, fiction or science texts
np_length	interval	Number of lexemes in complement NP (noun phrase)
np_distance	interval	Number of lexemes between the complement and the verb
head_noun_distance	interval	Number of lexemes between the head noun of the complement and the verb
np_case_proportion	interval	Proportion of complement phrase lexemes that are case-marked
animacy	categorical	Complement expresses a human, other animate, inanimate concrete or inanimate abstract referent

In addition, the model explored an interaction between animacy and pronouns, following a hypothesis that pronouns are more widespread with nominals with human referents.

Animacy is the only variable including NA values (in 578 observations out of the 9,866 observations in the dataset, including 137 in the canonical dataset and 441 in the oblique dataset). In these instances, the animacy status was unclear in the sentence, e.g. due to ambiguous pronouns.

The only random effect is the corpus file from which the clause originates (n = 602). Observations in the corpus are not entirely independent as several are likely generated by the same author. This may partially be captured by using files as random effect.

#### 4.4.2 Results of mixed logistic regression

The aim of the model is to assess, which features of usage distinguish between oblique and canonical argument verbs. The final model is as follows:

$$\text{verbtype} \sim \text{order} + \text{pronoun} + \text{number} + \text{tense} + \text{negation} + \text{register} + \text{np\_length} + \text{np\_distance} + \text{head\_noun\_distance} + \text{np\_case\_proportion} + \text{animacy} + \text{pronoun} * \text{animacy} + (1 | \text{file})$$

The output of the model is as follows:

**Table 9:** Results of logistic regression

Condition	Estimate ( $\beta$ )	Standard error	P value
Intercept	0.39668	0.13363	0.002992**
order = preverbal	0.07806	0.05411	0.149720
pronoun = yes	-1.89350	0.19202	< 2e-16***
number = plural	-0.19751	0.05424	0.000271***
tense = pastsimple	-0.23251	0.05106	5.12e-06***
tense = perfect	0.22773	0.09040	0.011767*
tense = pluperfect	0.14093	0.13594	0.299879
negation = yes	-0.06525	0.08326	0.433218
register = fiction	0.33777	0.07302	3.73e-06***
register = science	0.33955	0.07003	1.24e-06***
np = length	-0.11420	0.02207	2.29e-07***
np = distance	-0.00507	0.02002	0.800022
head noun distance	-0.02131	0.01484	0.151031
NP case proportion	-0.68745	0.11212	8.71e-10***
animacy = human	0.59595	0.08069	1.57e-13***
animacy = other animate	0.70442	0.27235	0.009697**
animacy = inanimate concrete	-0.43565	0.07115	9.20e-10***
pronoun = yes animacy = human	1.87174	0.21538	< 2e-16***
pronoun = yes animacy = other animate	2.38795	0.91926	0.009385**
pronoun = yes animacy = inanimate concrete	0.66292	0.41391	0.109247

Intercept depicts a hypothetical situation where each factor is represented by its most common value. This means that in Table 9, intercept describes a clause with a postverbal non-pronoun singular argument that refers to an inanimate abstract entity. The verb is in present tense and expresses affirmative mood. The clause originates from the journalism register. In such a context, oblique argument verbs are much more likely to be used than canonical object verbs ( $\beta = 0.397, p = 0.003$ ). Four tested factors have no effect on argument structure type: word order, negation, NP and head noun distance from the verb. The rest are discussed in the following sections.

## Argument phrase

Oblique and canonical argument verbs differ in terms of various argument phrase properties. The most striking finding in this analysis is that oblique verbs are more likely than canonical verbs to select arguments with human animate referents and less likely to select arguments with inanimate referents, especially inanimate concrete referents. Only 14.1% of the canonical dataset contains arguments with human referents, compared to 24.4% in the oblique dataset.

Although this is partially due to Experiencer and Recipient roles, the effect does not depend on them. These two roles account for less than half of the 991 human animate oblique arguments (38.5%). After removing Experiencers and Recipients from both dataset, the oblique argument dataset still includes a significantly higher proportion of human animate referents than the canonical object dataset (15% and 13.1%, respectively,  $\chi^2 = 6.2457$ ,  $df = 1$ ,  $p = 0.01$ ).

This means that oblique argument verbs are more likely than canonical argument verbs to be used for talking about events/states involving humans. Verbs with frequent human Themes include *andestama* ‘forgive’, *armuma* ‘fall in love with’, *hoolima* ‘care about’. Generally human Goals include *lasuma* ‘lie upon’ and human Sources include *hoovama* ‘emanate from’.

Regarding inanimate referents, canonical object verbs are more likely to occur with arguments that have concrete inanimate referents (17.9% vs 10.5%,  $\chi^2 = 99.608$ ,  $p < 0.0001$ ), but also with arguments that have inanimate abstract referents (67.5% vs 64.1%,  $\chi^2 = 11.611$ ,  $df = 1$ ,  $p = 0.0007$ ).

In addition to referring to humans more often, oblique arguments pronominalise differently from canonical objects. The interaction used in the model shows that pronominality heavily depends on the animacy of the complement. The model shows pronominalised arguments with inanimate abstract referents are more likely to occur as canonical objects than as oblique arguments, but pronominalised human arguments are more likely to occur with oblique case marking.

Hence, even though pronouns are generally more frequent in the oblique dataset than in canonical dataset ( $n = 999$ , 22.2% of the dataset, compared to 15.5% in the canonical dataset), it is mostly arguments with human referents that pronominalise — they account for 91.8% of all pronouns in the dataset. Arguments with inanimate referents are more likely to occur as pronouns in the canonical dataset. Oblique arguments are also more likely than canonical objects to occur in singular. Altogether, singular arguments make up 80.9% of the oblique dataset and 77.5% of the canonical dataset.

Finally, oblique arguments are generally shorter (in terms of lexeme count) than canonical objects (oblique: mean = 1.95 lexemes,  $sd = 1.54$ ; canonical: mean = 2.11,  $sd = 1.61$ ). This is likely due to the higher proportion of human referents among oblique arguments (mean length of human arguments = 1.37, other animate = 1.60, inanimate concrete = 1.79 and inanimate abstract = 2.39 lexemes). In addition, Table 9 shows an effect of the proportion of NP that is case-marked.

However, a distinction between how many of the NP lexemes bear case-marking only exists in intercept conditions (with inanimate abstract complements). Overall there is no effect.

## **Verb and clause**

The only distinction related to verb forms is that canonical object verbs are more likely to be in past simple (39.8% of their instances) than oblique argument verbs (34.8%). Both perfect (6.3% and 7.7%, respectively) and pluperfect (2.7% and 3.6%, respectively) are more frequent forms in oblique argument verbs. This will be further discussed in section 4.4.3, which addresses the tense differences between individual structures.

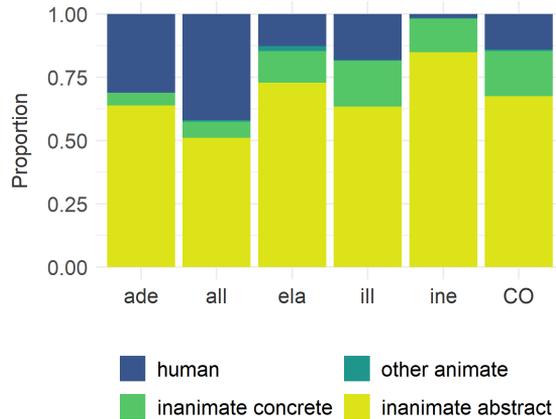
In which regards clause context, oblique argument verbs are more likely than canonical verbs to occur in science register (canonical: 34.4%, oblique: 38.5%) and fiction register (canonical: 32.3%, oblique: 34.6%), and less likely to occur in journalism (canonical: 33.3%, oblique: 26.9%).

### **4.4.3 Structure-based analysis**

The previous section described the ways in which the usage of the two verb types is distinct. However, this thesis also aims to investigate patterns distinguishing between oblique argument structures, outlining which are used most similarly to the canonical structure. This section will focus on the factors used in the regression analysis above from the perspective of individual argument structures. Structures' distinctions are summarised in the end.

## Complement

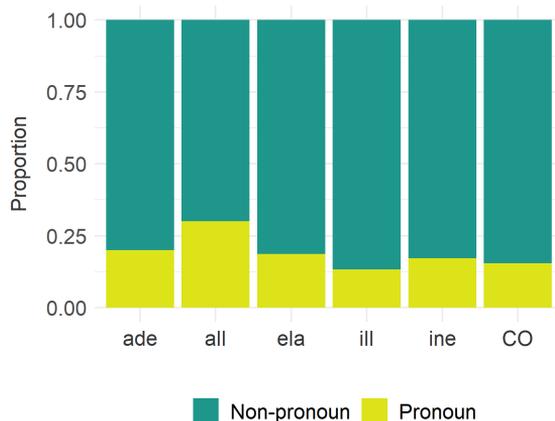
Argument structures with spatial cases were generally shown to differ from the canonical structure in terms argument referent animacy. In a structure-based perspective, however, it appears that relative and illative marked arguments have an referent animacy distribution similar to that of canonical objects (Fig. 10).



**Figure 10:** Animacy of the argument referent per argument structure (CO = canonical object in partitive, genitive or nominative).

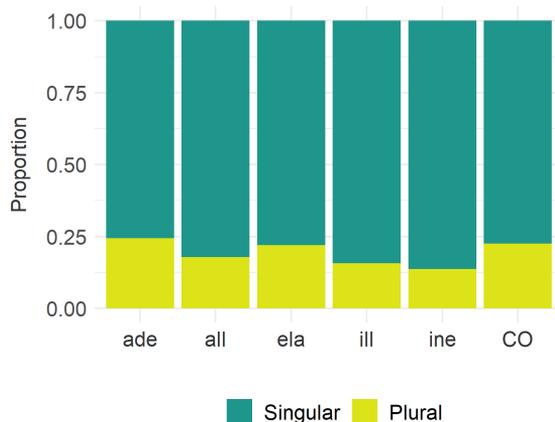
Similarly to illative arguments, relative arguments and canonical objects do not statistically differ in terms of the proportion of human animate referents (relative arguments vs canonical objects: 12.7% and 14%, respectively). They do, however, refer to inanimate concrete entities significantly less frequently (relative arguments vs canonical objects: 12.5% vs 17.9%,  $\chi^2 = 23.153$ ,  $df = 1$ ,  $p < 0.0001$ ). Allative and adessive arguments have a high proportion of human animates, while inessive arguments are dominated by inanimate abstract referents.

Most spatial case argument structures also differ from the canonical structure in terms of pronouns, marking them more often. The effect is the strongest with allative ( $\chi^2 = 174.08$ ,  $p < 0.0001$ ). Adessive and inessive structures, however, include pronominal arguments as frequently as the canonical structure with no statistically significant distinction (Fig. 11).



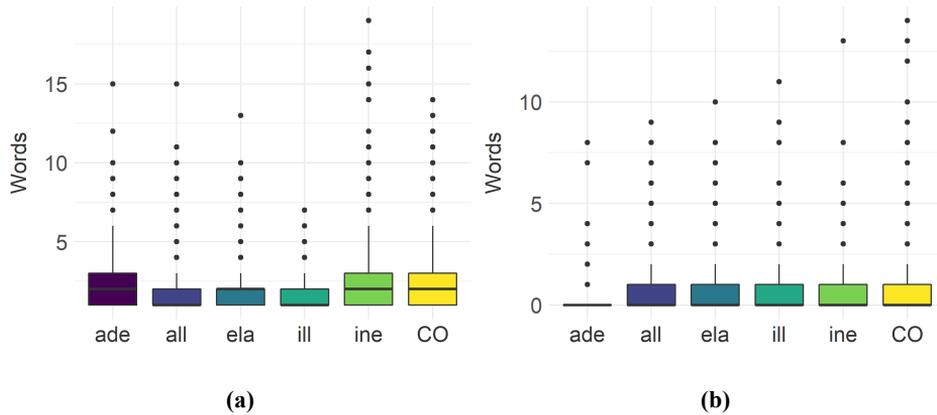
**Figure 11:** Pronoun proportions per argument structure (CO = canonical object in partitive, genitive or nominative).

Structures also vary in terms of number. No significant distinction exists between the number proportions of adessive and canonical structures, nor between elative and canonical structures. The other three spatial case structures are significantly more likely to include singular arguments than the canonical structure (Fig. 12).



**Figure 12:** Number proportions per argument structure (CO = canonical object in partitive, genitive or nominative).

In terms of mean length of NP structure (in lexeme count), elative arguments (mean = 1.94 lexemes) are the closest to canonical objects (mean = 2.11 lexemes, see Fig. 13a).

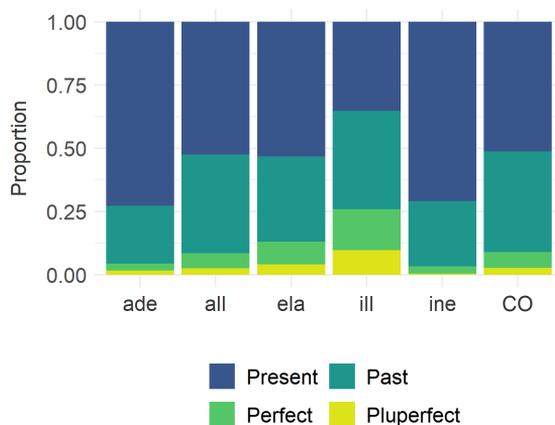


**Figure 13:** NP length (a) and distance from the verb (b) in lexemes (CO = canonical object in partitive, genitive or nominative).

In addition, as seen in Fig. 13b, the distance between the argument and the verb (in lexeme count) is rather similar in most argument structures with no statistically significant differences. However, adessive arguments tend to occur significantly closer to the verb than canonical objects (0.44 vs 0.8 lexemes between the argument phrase and the verb). In addition, adessive and elative arguments are the only groups not significantly different from canonical objects in terms of the proportion of lexemes in the complement phrase which are case marked (adessive: 75.4%, elative: 79.5%, canonical objects: 79.6%).

## Verb

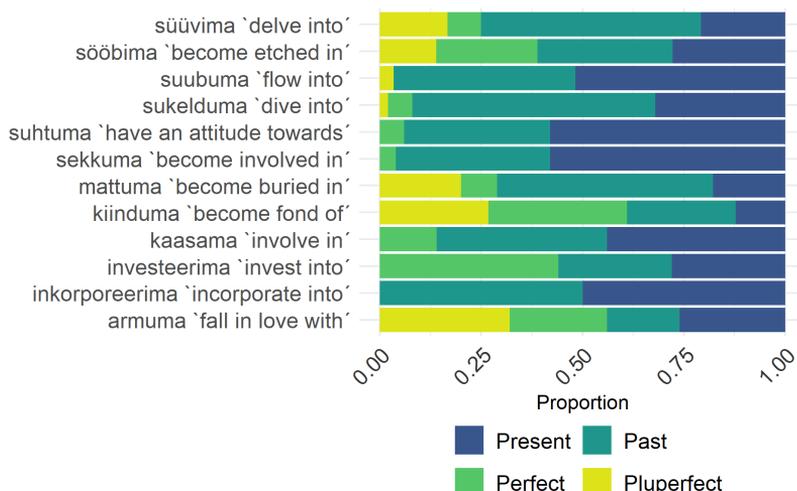
Spatial cases vary greatly in terms of tense patterns. The only oblique argument verbs which show no significant distinction from the present tense proportions of canonical object verbs are allative and elative (Fig. 14):



**Figure 14:** Tense proportions per argument structure (CO = canonical object in partitive, genitive or nominative).

Allative-selecting verbs also occur in past simple as frequently as canonical object verbs. Verbs selecting adessive and inessive as distinct from verbs in other structures in that they occur in past simple less frequently. Regarding adessive, this is almost entirely due to three near-synonymous verbs selecting Themes: *baseeruma*, *rajanema* and *pōhinema* (‘be based on’), which predominantly occur in present simple. All three were coded as K-states in section 4.2.2, expressing entirely abstract and static properties.

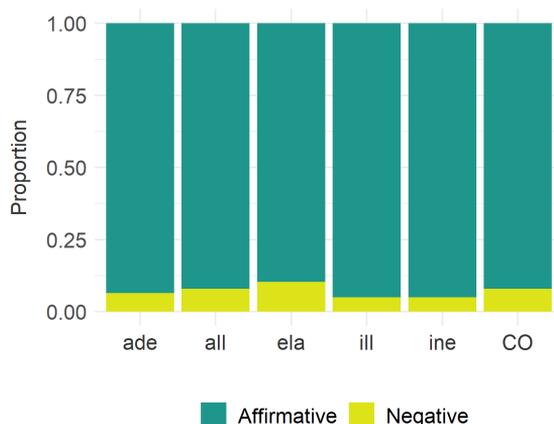
Tense patterns show an interesting aspect about the verbs selecting illative. Illative argument structure differs from other spatial case argument structures in that many of its verbs frequently occur in aspectual tenses, i.e. perfect and pluperfect. As seen in Fig. 15, half of all verbs selecting illative arguments occur in aspectual tenses in at least 25% of their instances:



**Figure 15:** Tense in verbs selecting illative arguments.

These six verbs include *armuma* ‘fall in love with’ (56%), *investeerima* ‘invest in’ (44%), *kiinduma* ‘become fond of’ (61%), *mattuma* ‘become buried in’ (29%), *sööbima* ‘become etched in’ (39%) and *süüvima* ‘delve into’ (25%). All are non-states, but almost all nevertheless describe non-volitional events. This means that there are other ways in which oblique argument verbs may be stative, besides lexical stativity. Instead of being used for referring to the process of falling in love or becoming buried in something, they appear to be frequently used to refer to the end state of that process, expressed with the perfect form.

Finally, negation patterns are rather similar across all structures (Fig. 16):

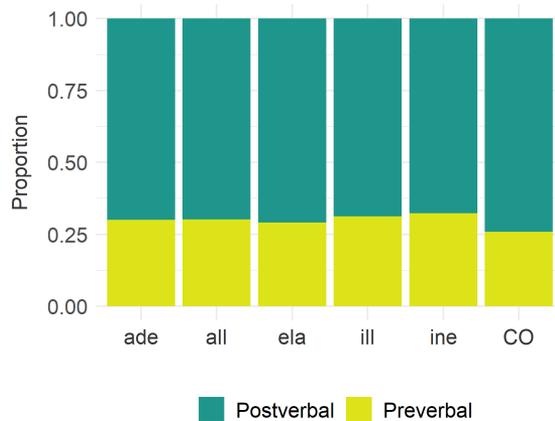


**Figure 16:** Negation proportions per argument structure (CO = canonical object in partitive, genitive or nominative).

Adessive and allative argument verbs are negated as frequently as canonical object verbs, showing no statistical distinction. Other spatial case structures vary in terms of mood. Elative-selecting verbs are negated significantly more often than canonical object verbs ( $\chi^2 = 8.6492$ ,  $df = 1$ ,  $p = 0.003$ ) while verbs selecting illative and inessive arguments occur in negation less often than canonical object verbs (verbs selecting illative vs canonical object verbs:  $\chi^2 = 5.2148$ ,  $df = 1$ ,  $p = 0.02$ ; verbs selecting inessive vs canonical object verbs:  $\chi^2 = 4.8545$ ,  $df = 1$ ,  $p = 0.03$ ).

### Clause

Word order does not vary greatly between the six argument structures, preverbal arguments ranging from 24.8% to 30.6%:

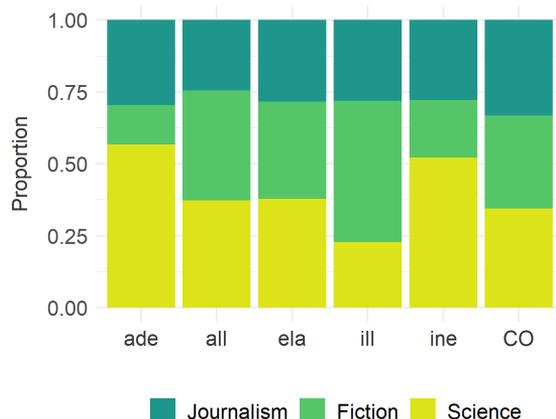


**Figure 17:** Word order per argument structure (CO = canonical object in partitive, genitive or nominative).

In most oblique structures arguments occur preverbally significantly more frequently than canonical objects (verbs selecting elative vs canonical object verbs:  $\chi^2 = 6.4502$ ,  $p = 0.01$ )<sup>29</sup>, see Fig. 17.

In addition, verbs selecting canonical objects exhibit an almost perfectly even distribution between the three registers (33.2% in journalism, 32.3% in fiction and 34.4% in science), as seen in Fig. 18:

<sup>29</sup> There is no statistically significant distinction between the word order of adessive selecting verbs and canonical object verbs, most likely due to the small sample of clauses with adessive selecting verbs.



**Figure 18:** Register proportions per argument structure (CO = canonical object in partitive, genitive or nominative).

Oblique structures with the most similar register distribution include the allative structure (24.5% in journalism, 38.2% in fiction and 37.3% in science) and the elative structure (28.3%, 33.9% and 37.8%, respectively). Verbs selecting adessive and inessive arguments mostly occur in science texts, while illative-selecting verbs frequently occur in fiction texts.

### Summary of case-based analysis

Table 10 summarises the lack of statistical distinctions between oblique and canonical argument structures in section 4.4.3:

**Table 10:** A usage factor comparison between the canonical structure and each individual oblique structure.

Factor	Oblique structures most similar to the canonical structure
Word order	Adessive
Pronoun	Adessive and inessive
Number	Adessive and elative
Tense	Allative and elative
Negation	Adessive and allative
Register	Allative and elative
NP length	Elative
NP distance from the verb	Allative, illative, elative and inessive
NP case marking	Adessive and elative
Animacy	Elative and illative

When the contents of Table 10 are summarised, elative argument structure is used most similarly to the canonical object structure, with no statistically significant distinction in seven out of ten factors, followed by adessive (five factors), allative (four factors), inessive (two factors) and illative (two factors).

## 4.5 Summary

This chapter aimed to answer three research questions:

RQ1: Are oblique argument structures assigned to lexically less transitive verbs than canonical argument structures? If yes, in what way are these verbs less transitive?

RQ2: To what degree can spatial semantics be detected in Estonian verbs occurring in oblique argument structures and how does this compare to canonical object verbs?

RQ3: Does the usage of verbs with oblique argument structures differ from that of canonical argument structures? If yes, in what way?

Study 1 gives RQ1 a rather clear answer — almost all oblique argument structures are indeed much more likely to be assigned to verbs expressing K-states (verbs expressing properties and relationships, Vendlerian states), but they are also more likely to occur with verbs expressing non-dynamic events (*viitama* ‘refer to’).

The two types of structures are similar in that the verbs with which they occur have similar lexical aspect patterns. A link between atelicity and oblique verbs was assumed not only because stativity entails lexical atelicity, but also because certain argument realisation approaches regard lexical aspect as playing a central role in the way semantics maps onto syntax in terms of event participant marking (Tenny, 1992). In Estonian, however, a verb’s lexical aspect appears to not play a role in determining whether a verb is assigned an oblique or a canonical case frame. However, as discussed in chapter 3, aspect is central to the Estonian system of differential object marking.

It is also interesting that experiencer (and possessor) verbs constitute a verb type that occurs with almost all spatial case argument structures more frequently than with the canonical structure, rather than merely being limited to adessive and allative structures. Hence, it is not only the Experiencer role that makes such verbs likely to acquire oblique structures, but the event/state in general. Elative and illative arguments occur in experiencer verbs almost as frequently as adessive and allative arguments. These verbs simply assign nominative to the Experiencer and mark the Stimulus/Theme non-canonically (*hoolima* ‘care about’). This suggests that oblique argument structures are linked to the lexical semantics of verbs in a more broad way, rather than there being an exclusive link between a few semantic cases and a particular semantic roles. This observation is easily missed in

an example-based study taking a non-quantitative approach, because strong pre-conceptions exist about the link between allative/adessive and the Experiencer role.

Results also show both oblique and structure-based effects, as was hypothesised in the overview of the study presented in section 3.7. As a structure type effect, increased stativity describes almost all oblique argument structures more so than the canonical structure. As a structure-based effect, this does not appear to apply to the illative structure. By the measures of the present study, verbs selecting illative arguments do not appear to be lexically less transitive than verbs selecting canonical objects. However, as was shown in section 4.4.3, decreased transitivity may be expressed in illative verbs by way of usage patterns rather than their lexical semantics. Namely, half of the illative verbs were shown to occur unusually frequently in perfect and pluperfect tense, which mark telic aspect. Hence, one may postulate a general link between oblique structures and stativity, which may be expressed lexically (verbs selecting elative, allative, adessive and inessive) as well as via morphosyntax (verbs selecting illative).

Study 2 answered RQ2. While oblique argument verbs code spatial roles more frequently than canonical object verbs, it is nevertheless expressed by a minority of verbs in that group (less than a third). Most oblique verbs describe events with participants in semantically bleached roles where the case only marks the fact that the nominal expresses a participant to the event described by the verb. While the study may not conclude that the majority of *all* oblique argument verbs select cases with entirely bleached meanings, it does show that there is no tendency for oblique argument verbs to always or frequently select cases, the original semantics of which remain present in the clause.

Finally, in order to answer RQ3, study 3 explored whether various usage factors differ between the two verb groups. It found that oblique arguments are generally more likely to refer to human participants than canonical objects (even without Experiencer and Recipient roles), but that this effect pertains to adessive and allative structures. This is also likely behind the observation that oblique arguments are generally shorter (in lexeme count) than canonical objects, and more often in singular. Oblique arguments not referring to humans are less likely than canonical objects to occur as pronouns.

In addition, as oblique verbs are more stative, they occur in present tense more often and in past simple less often than canonical object verbs. They are also more frequent in science and fiction genres, and less frequent in journalism.

### 4.5.1 Discussion

The present results have interesting implications on the way oblique verbs are conceptualised. Increased stativity in events/states means that in many contexts it is difficult to clearly distinguish between, for instance, the nominative and the spatial case argument in a clause in terms of hierarchy. If the transitivity of a two-place verb is low, both participant roles are rather abstract and bleached of their Agent and Patient properties, or Proto-Agent and Proto-Patient entailments (Dowty, 1991). This is because these properties are interlinked with the lexical transitivity of verbs. Hence, when focusing on state verbs with close to no transitivity, where the verb merely marks an abstract relationship between two participants, there is very little semantics to postulate a hierarchical distinction between the two arguments of the verb. This is reflected in the higher proportion of human animate referents among oblique arguments, compared to canonical objects. In oblique argument structures, the more animate and more agentive participant might be marked non-canonically (e.g. the allative Experiencer of *meeldima* ‘like’), but it might also occur as a canonical nominative subject (*hoolima* ‘care about’). All in all, the fact that oblique argument structures show decreased transitivity affects all aspects of the way we ought to talk about oblique arguments. For instance, because prototypical subject and object properties are less apparent among these verbs (and often missing), a more general term ‘oblique argument’ might be more accurate than ‘oblique object’ or ‘oblique subject’.

Subject and object labels are not the only problematic terminology when talking about oblique verbs. The fewer Proto-Agent and Proto-Patient entailments are represented, the more difficult it is to assign semantic roles. For instance, *eelnema* ‘precede’ is a relationship between two often inanimate entities, where neither is volitional nor agentive. No roles fit well. When both are regarded as Themes, roles become unhelpful and misleading, because they leave the impression that both roles are analogous, even though one is clearly more in focus than the other. Hence, the set of roles might benefit from more research on stative verbs, because these participants, too, form semantic patterns in relation to each other and the verb.

Many new questions arise from these studies. First, while a general effect of stativity was observed, structure-based distinctions were observed as well. Is there a connection between elative being bleached in terms of semantic roles and the elative structure being used in a way most similar to the canonical structure? In order to explain the observed structure-based differences, chapter 5 will focus on spatial cases in general, investigating their individual degrees of grammaticalisation and illustrating how both a synchronic and a diachronic perspective are needed to tell the whole story on argument structures.

Second, the effect of stativity brings us back to argument status. What are the implications of decreased transitivity on the quality of the cognitive connection between arguments and verbs? Do states encode arguments of similar strength to those of transitive verbs? What role does morphosyntax play in this?

In any case, both types of questions outlined above require wholly different perspectives and methods than the ones answered in the present chapter. In order to address these questions, case grammaticalisation is investigated in chapter 5 and the argument status of oblique arguments is explored in chapter 6.

## 5 CASES AND ARGUMENT STRUCTURES

### 5.1 Introduction

This chapter investigates the second fundamental variable in terms of which argument structures vary: the case selected by the verb. The three studies presented in this chapter take a diachronic perspective to case functionality, approaching it by synchronic means (Aigro and Vihman, 2023).

Many studies refer to a fundamental distinction between structural and semantic cases (Blake, 2001) (see section 2.3). Broadly speaking, the function of structural case is viewed as encoding arguments and distinguishing between them, e.g. case helps to formally differentiate between Agents and Patients in many two-place verbs in Estonian. Semantic cases, however, are assumed to mark semantically more specific roles (Wunderlich and Lakämper, 2001; Blake, 2001; Lestrade, 2010), e.g. Time or Manner.

However, as was shown both in section 2.3, the two case types are not discrete categories, but parts of a diachronic grammaticalisation cline instead (Lestrade, 2010). The cline describes how lexical items (nominals and verbs) tend to grammaticalise into adpositions, which may become semantic case affixes, which in their turn have been noted to grammaticalise into structural case affixes. This implies that while some cases may be entirely semantic and others entirely structural, some cases may also fall in between, displaying properties of both types. Furthermore, Nichols (1983) posits that the phenomenon of multifunctional cases is a typological notion. Some languages have a great deal of functional overlap in their case systems, with most cases marking both adverbials and arguments (e.g. Finnish and Russian in Nichols's analysis), while other languages have clear structural and semantic case classes with virtually no overlap (e.g. Chechen-Ingush).

Estonian appears to be a language of the first type. Not only are a wide range of its semantic cases used both semantically and structurally, marking adverbials as well as arguments, but even its structural cases are regularly used for conveying meaning that is more specific than their role of merely marking event participation in arguments (e.g. telicity in DSM and DOM). This overlap in Estonian cases' functionality is the result of the diachronic process of grammaticalisation. During grammaticalisation, semantic cases become semantically bleached, meaning they become used with less cohesive semantics in a wider range of contexts (i.e. more similar to structural cases as parts of argument structures). We may assume that when they reach a certain level, they gain a new function where they occur in a particular syntactic configuration with particular types of verbs (see chapter 4), marking the participants of the events expressed by these verbs.

Hence, when focusing on semantic case in structural function, there is no way around the diachronic perspective of grammaticalisation, because it is the degrees of grammaticalisation of cases that fundamentally describe, which oblique argument structures are used in a language, and how. Hence, in order to obtain a

thorough picture of a language's argument structure system, its inner dynamics and labour division, both the diachronic and synchronic perspective are required.

The present chapter aims to show a link between the cases' degrees of grammaticalisation and various other properties of argument structures. It hypothesises that the spatial cases, the argument structures of which chapter 4 found to be 1) frequently used for marking spatial semantics in argument structure (section 4.3) and 2) used more similarly to the canonical structure (section 4.4), will also emerge as the most grammaticalised in the present chapter. The present chapter is also informative in the context of the argument status investigation in chapter 6, because it is unclear how cases' grammaticalisation degrees are linked to the way argument structures with said cases encode argument status.

In addition, the present chapter addresses another gap in what we know about the assignment of argument structures to verbs. Chapter 4 showed that verbs describing stative events and states are more susceptible to being assigned oblique structures, but also that this affects most oblique structures in a similar way. It also showed that many verbs do not use argument structures with spatial cases to mark spatial semantics. The question therefore remains, when various oblique argument structures lack spatial semantics and are assigned to similar types of verbs, how do they differ in terms of the number of verbs to which they have been assigned and what underlies this distinction? The thesis hypothesises that assignment as argument structures also depends on the grammaticalisation degrees of spatial cases.

This chapter focuses on six Estonian spatial cases, which have all been noted to mark oblique arguments. First and foremost, this chapter aims to investigate and compare their degrees of grammaticalisation. As there are no corpora old enough to conduct a diachronic investigation into most Estonian cases, this property will be assessed by means of three different synchronic variables: functional profiles (section 5.1.1), lexical variance (section 5.1.2) and syntactic productivity (section 5.1.3). All three are investigated by means of corpus studies. The following sections will outline each variable and its relevance to the question at hand.

### 5.1.1 Functional profiles

This thesis uses cases' functional profiles to investigate their grammaticalisation. As was discussed in sections 3.2.2 and 3.5, each Estonian spatial case has a number of functions for which it is used (see also Vainik, 1995). This bundle may be represented with a functional profile, describing the functions and their frequency in the usage of that case affix. For instance, adessive ('on') may mark physical locations (*laua-l* table-ALL 'on the table'), time (*laupäeva-l* Saturday-ALL 'on Saturday') and a bleached function, marking nothing but event participation (the argument of *baseeruma* 'be based on'). Some of these functions constitute semantically coherent sets (e.g. the time of events) while others are less coherent

and more bleached event/state participants. Some functions might be metaphorically connected to their original spatial function (the time of events), while others, at least in an intuitive perspective, are not related to the original meaning of the case (e.g. the elative topic argument in *raamat muusika-st* book.NOM music-ELA ‘a book about music’). Hence, the profiles of different cases vary in terms of the frequency of different types of functions as well as the total number of functions.

Functions are relevant to grammaticalisation, because the whole phenomenon essentially describes a functional change, designed as a tool for describing changes in lexical items along the lexical-functional (lexical-grammatical) scale. Becoming more grammatical entails several predictable changes in an item, e.g. becoming semantically more bleached (being used with less coherent semantics in more functions), phonologically shorter and syntactically more restricted (Heine et al., 1991). Hence, for a case, a highly varied functional profile indicates a high degree of grammaticalisation, as it contrasts with the profile containing a small number of semantically highly cohesive functions of less grammaticalised cases.

Cross-linguistic research gives us reason to assume asymmetry between the functional profiles of spatial cases. Some cases have been noted to be systematically more multifunctional than others. Rice and Kabata (2007) show that Goal-marking morphemes (‘allatives’) are polysemous cross-linguistically. They find a total of 33 senses, the most common of which are SPATIAL, PURPOSE, CONCEPTUAL (‘think about’), RECIPIENT and TIME (see also Haspelmath, 1989; Heine and Kuteva, 2002; Västi, 2011; Ylikoski, 2011).

In Estonian, elative is highlighted as highly multifunctional (Nurka, 2014; Viht and Habicht, 2019). Bleached functions have also been noted for adessive and allative (Matsumura, 1994; Lindström and Vihman, 2017), although they may be expected to be used for fewer functions than elative (section 3.2.2). All spatial cases, however, were shown to mark bleached roles that signify nothing but event participation at least to some degree in chapter 4. It is curious that this aspect, i.e. the degree to which a semantic case is used non-semantically is often not in the focus of function studies. When functions are investigated, the angle is often to demonstrate variance and analogical use, rather than use this variance to quantify degrees of bleachedness and structural use, describing a case in a wider perspective.

Currently there are no quantitative studies of Estonian cases which compare the degrees to which their uses are bleached. This type of information is, however, crucial for describing an argument structure system, because function frequencies underlie the traditional approach to case categories. “[...] all case-forms share both kinds of functions, but the syntactical function is *primary* with “grammatical” and *secondary* with “concrete” cases [...]” (Kuryłowicz, 1964, 179). This lays the foundation for a frequency-based approach to case and argument structure functionality.

## 5.1.2 Lexical selection patterns

In addition to the cases' functional profiles, grammaticalisation will also be explored by means of their lexical selection patterns. Broadly speaking, this variable describes the variety of lexemes a case marks. Although case affixes constitute inflectional morphemes and are therefore expected to be *able* to occur with all case-inflecting lexemes (nominals in Estonian), they tend to not be used with all lexemes. Instead, cases may be expected to have distinct patterns in terms of the variety of lexemes they mark, that are informative of the overall grammaticalisation degrees of these cases.

The lexical semantics of a nominal often sets restrictions on the set of cases it commonly occurs. Karlsson (1986) points out that while nominal paradigms can be considered to be full in an ideal sense (except for rare instances of defective slots, mostly occurring as *pluralia* and *singularia tanta*, e.g. the lack of singular for 'scissors'), they are still governed by semantic restrictions which affect their overall case distribution. For instance, many mass nouns, abstract words and proper nouns have very infrequent plural forms (Karlsson, 1986, 22). Hence, a higher number of lexemes occurring with a case indicates a wider range of different contexts for that case affix. Wider range of contexts, in its turn, indicates higher functionality in case usage and therefore a higher degree of grammaticalisation (Aristar, 1997; Lestrade, 2010).

Silverstein (1976) and Aristar (1997) discuss these distinctions in the lexical patterns of cases. They show that both various cases have particular types of NPs which they tend to mark (with which they are 'congruent'). Broadly speaking, there are high-animacy cases (e.g. dative) and low-animacy cases (e.g. locative, instrumental) (Aristar, 1997). The type-congruent nominals of spatial cases are assumed to have concrete referents ('house'). However, it may nevertheless happen that a spatial case is used for marking type-incongruent nominals (e.g. nominals with human referents). When this happens, the entire phrase is commonly reinterpreted. Hence, the adessive ('on') form of 'mum' is not used for marking a spatial relation (something being on top of mum's head), but instead a more abstract relation between mum and some event or state. Such uses may become more frequent over time, which will eventually lead to a change in that case's grammaticalisation more generally (Aristar, 1997, 318). This change is not only evident in semantics. Aristar (1997) points out that the multifunctionality of the spatial 'case-complex' (a multifunctional case) *-la* in Yidiny (Pama-Nyungan, Australia) is also reflected in its morphology, because the infix *-n-* is added to the *-la* suffix when the case marks nominals with human referents.

This notion is central to the present thesis. Extending the usage of a case to type-incongruent nominals will lead to a stage in its grammaticalisation where the semantics of a case-marked nominal is not simply the sum of the meaning of its stem and case affix. Instead, the extension renders a different, 'marked' interpretation, which can start to affect the further use of that case, possibly causing

it eventually to grammaticalise further. Aristar (1997) calls these multifunctional affixes ‘case-complexes’, given that they render different meanings in different contexts.

All in all, lexical variance is directly connected to case functionality and grammaticalisation. It therefore constitutes a productive measure with which to assess degrees of grammaticalisation. This thesis uses several measures used in morphological productivity research to assess this property of cases. This approach is somewhat connected to the Behaviour Profile methodology (Gries and Divjak, 2009). Janda and Lyashevskaya (2011) use this framework to derive semantic verb types from inflection distributions. These distributions are viewed as a strongly empirical and statistical approach that may benefit many branches of linguistics and allow various theoretical hypotheses to be empirically confirmed by means of behaviour patterns (Eckhoff and Janda, 2014).

### **5.1.3 Syntactic productivity**

The third study in this chapter will address the grammaticalisation of spatial cases by focusing on the productivity of the argument structures in which they occur. Namely, various canonical and non-canonical argument structures are expected to vary in terms of their syntactic productivity (Barðdal, 2008).

Syntactic productivity describes the availability of individual argument structures for marking novel verbs Barðdal (2008). The concept is analogous to morphological productivity, where, for instance, derivational morphemes are viewed as recurrent patterns with different degrees of ‘availability’ to variable parts, i.e. individual lexemes. In syntactic productivity, argument structures constitute the patterns while verbs constitute the variable parts. Argument structures used with a wider range of verbs are said to be more productive, meaning that speakers are more likely to use them with novel verbs. Following the discussion on lexical variance above, one may assume more grammaticalised cases to render more productive argument structures.

In the system proposed by Barðdal (2008), there are two ways in which argument structures can be productive. They can constitute productive patterns (high-level extensions) or analogy-based patterns (low-level extensions) (Bauer, 2001, productivity and creativity in).

An argument structure may be regarded as a productive pattern if it occurs with a wide range of verbs. Its overall meaning is generalised to the point that it may be used in many different contexts and with a variety of verbal lexical semantics. Using such a structure on novel verbs would be considered a high-level extension, meaning that its availability to novel verb lemmas derives from the diversity of its current verb types and their non-specific semantics. This means that a high-level extension argument structure is so broadly used that speakers assess it as highly available to novel verbs (Barðdal, 2008).

However, if an argument structure does not occur with enough verbs to be considered a productive pattern, it may still become extended to novel verbs. It can happen based on the semantic analogy between verbs, meaning that the structure would only be available to verbs that are somewhat synonymous to its current verbs, with which it is already used. Furthermore, Barðdal (2008) argues that in such instances, token count matters. If several such argument structures are in competition, both marking relatively similar semantics, speakers are more likely to extend the argument structure where existing verbs have a higher overall token count (Barðdal, 2008). In any case, analogy-based low-level extension patterns are not to be viewed as inferior to productive high-level patterns. Barðdal (2008) shows that novel verbs in Icelandic are often assigned argument structures based on semantic analogy. This means that when speakers have encountered a verb for the first time, they have assigned it the argument structure used with existing verbs that are semantically analogous to the novel verb (Barðdal, 2008, 88).

All in all, syntactic productivity constitutes a useful variable for assessing case grammaticalisation. It will also provide an informative outline of various argument structures and their prominence in Estonian.

#### **5.1.4 Research questions**

All three variables — functional profiles, lexical selection patterns and syntactic productivity — serve the common aim of outlining the degrees of grammaticalisation of Estonian spatial cases. Each is explored in a separate study in the present chapter. Hence, the chapter altogether has three research questions:

RQ4: How do spatial cases vary in terms of their functions and which spatial cases are used in the most multifunctional and semantically bleached way?

RQ5: Which spatial cases have the widest lexical selection patterns?

RQ6: Which oblique argument structures are the most productive?

Following the results presented in section 4.3 and the background on Estonian spatial case functionality (section 3.5), this chapter hypothesises that elative and allative constitute the most grammaticalised cases as per these variables.

Study 1 (RQ4) will be presented in section 5.2. Section 5.3 will outline study 2 (RQ5) and section 5.4 will present study 3 (RQ6). Finally, section 5.5 will discuss the results of all three studies.

## 5.2 Study 1: Functional profiles of cases

This study is designed to answer RQ1 — “How do spatial cases vary in terms of their functions and which spatial cases are used in the most multifunctional and semantically bleached way?” For this, it explores the functions of Estonian spatial case affixes by means of a balanced corpus sample, in which functions were manually coded. Section 5.2.1 describes the data and method of this study while section 5.2.2 presents its results.

### 5.2.1 Data and method

200 instances were randomly drawn from the Balanced Corpus of Estonian for each spatial case (allative, adessive, ablative, illative, inessive, elative). No restrictions were set for the sample as it was intended to be representative of all uses of these cases. The sample is freely available (Aigro, 2022b).

Each instance was extracted together with the clause in which it occurred. Contextual information was used to code the function of the case. Instead of using existing function lists of spatial cases (see section 3.2.2), the levels used for coding were derived from the data in an *ad hoc* manner. As there is no quantitative overview of the functions of most Estonian spatial cases, this method was deemed more empirically grounded, allowing for better comparison between cases. In addition, the aim was to keep the number of functions as small as possible to reduce subjectivity in coding and allow for maximal comparability between cases with slightly different semantics, while keeping as many as necessary to capture the main distinct functions of spatial cases.

Table 11 lists the functions found for each case:

**Table 11:** Functions found for each case affix in a sample of 200 tokens per case.

Illative ‘into’	Inessive ‘in’	Elicative ‘out of’
CONCRETE LOCATION	CONCRETE LOCATION	CONCRETE LOCATION
ABSTRACT LOCATION	ABSTRACT LOCATION	ABSTRACT LOCATION
SHAPE	MANNER	REASON
TIME	TIME	COMPARISON
THEME	PROPERTY	PROPERTY
	THEME	TIME
		SUBJECT
		PARTITIVE
		TOPIC
		THEME
Allative ‘onto’	Adessive ‘on’	Ablative ‘from on’
CONCRETE LOCATION	CONCRETE LOCATION	CONCRETE LOCATION
ABSTRACT LOCATION	ABSTRACT LOCATION	ABSTRACT LOCATION
RECIPIENT/BENEFICIARY	EXPERIENCER/POSSESSOR	SOURCE
EXPERIENCER	MANNER	MANNER
THEME	REASON	PROPERTY
	TIME	TIME
	THEME	THEME

### Illative functions

CONCRETE LOCATION refers to physical locations and ABSTRACT LOCATION to non-physical locations (*hinge* soul.ILL ‘into the soul’). Illative also marks TIME (51a) and SHAPE (51b), which is somewhat similar to ABSTRACT LOCATION.<sup>30</sup>

- (51) a. Vanim torn kuulu-b seitsmenda-sse sajandi-sse  
oldest tower.NOM belong-PRS.3SG seventh-ILL century-ILL  
‘The oldest tower belongs in the seventh century’
- b. Müür on ratta-sse painuta-tud  
wall.NOM COP.PRS.3SG wheel-ILL bend-PPP  
‘The wall is bent into a wheel’

Illative THEME occurs with *surema* ‘die of’ and *puutuma* ‘have to do with’, but it also marks nominal arguments (*usaldus süsteemi* trust.NOM system.ILL ‘trust in the system’).

<sup>30</sup> It was nevertheless coded as a separate function, because it answers the question *kuidas?* ‘how?’ rather than *kuhu?* ‘into what?’/‘where?’.

## Inessive functions

CONCRETE LOCATION includes *apteegi-s* pharmacy-INE ‘in the pharmacy’ while ABSTRACT LOCATION includes *novelli-s* short.story-INE ‘in the short story’. Inessive was also found to mark MANNER (52a) and PROPERTY (52b):

- (52) a. Nad ütle-sid se-da koori-s  
3PL.NOM say-PST.3SG this-PAR choir-INE  
‘They said it in chorus’  
b. Pealinn on ärevuse-s  
capital.NOM COP.PRS.3SG anxiety-INE  
‘The capital is anxious (in anxiety)’

Inessive TIME includes *jaanuari-s* january-INE ‘in January’. THEME only includes verbal arguments, e.g. those of *veenduma* ‘become convinced of’, *kahtlustama* ‘suspect someone in’, *osalema* ‘take part in’ and *süüdistama* ‘accuse someone in’.

## Elicative functions

Spatial functions include CONCRETE LOCATION (*koti-st* bag-ELA ‘out of the bag’) and ABSTRACT LOCATION (*müügi-st* sale-ELA ‘from the sale’). COMPARISON function is seen in (53a), PARTITIVE in (53b), REASON in (53c) and PROPERTY in (53d).

- (53) a. Laps on keskmise-st suurem  
child.NOM COP.PRS.3SG average-ELA larger  
‘The kid is larger than average’  
b. Pool ressursi-st on kasutamata  
half.NOM resource-ELA COP.PRS.3SG unused  
‘Half of the resources have not been spent’  
c. Keha on istumise-st kange  
body.NOM COP.PRS.3SG sitting-ELA sore  
‘Body is sore from sitting’  
d. Meeskonna prantslase-st liider  
team.GEN French.person-ELA leader.NOM  
‘The team’s French leader’

Elicative-marked TIME commonly refers to the start date of things, e.g. *uue-st aasta-st* new-ELA year-ELA ‘Starting from new year’. Arguments coded as having the function TOPIC may be selected by both nominals and verbs, as in (54):

- (54) Kuul-si-me skandaali-st alles hiljem  
hear-PST-1PL scandal-ELA only later  
‘We only heard about the scandal later’

Finally, relative can mark semantic subjects in passives (SUBJECT function), which are somewhat similar to REASON, as in (55):

- (55) Naine on kuumuse-st piina-tud  
 woman.NOM COP.PRS.3SG heat-ELA torture-PPP  
 ‘The woman is tortured by/from the heat’

Differently from REASON, however, the relative phrase in (55) may be easily substituted with a *poolt* (‘by’) phrase, which canonically marks demoted subjects in passives. REASON, on the other hand, occurs with an adjective (53c) and therefore resists alternation with the *poolt* phrase. Finally, most relative THEME instances constitute verbal arguments (e.g. the arguments of *unistama* ‘dream of’ and *aru saama* ‘understand’)

### Allative functions

CONCRETE LOCATION describes contexts such as *seina-le* wall-ALL ‘on the wall’ while ABSTRACT LOCATION includes *töö-le* work-ALL ‘to work’.

The RECIPIENT and BENEFICIARY functions share much overlap in the use of Estonian allative, which is why they were kept together under a single label, named RECIPIENT/BENEFICIARY. For instance, (56) is ambiguous between the two:

- (56) Laste-le on ravim-id tasuta  
 child.PL-ALL COP.PRS.3SG medication-NOM.PL free  
 ‘Medications are free of charge for the children’

The EXPERIENCER function includes non-volitional, often animate arguments, e.g. those of *pähe tulema* ‘think of’, *selguma* ‘become clear to’, *paistma* ‘seem’, *tunduma* ‘appear to’. Allative THEME function occurs with arguments entirely bleached of spatial semantics, or cohesive semantics of any type, e.g. the argument of *reageerima* ‘react to’.

### Adessive functions

CONCRETE LOCATION includes *Šotimaa-l* Scotland-ADE ‘in Scotland’, ABSTRACT LOCATION includes *teadvuse-l* consciousness-ADE ‘conscious’. Allative EXPERIENCER function constituted a fairly straightforward group, but adessive EXPERIENCER is often difficult to distinguish from POSSESSOR, e.g. (57), where pediatricians are both Possessors and Experiencers:

- (57) Pediaatrite-l on keelelise-d raskuse-d  
 pediatrician.PL-ADE COP.PRS.3SG linguistic-NOM.PL difficulty-NOM.PL  
 ‘Pediatricians have linguistic difficulties’

Hence the label EXPERIENCER/POSSESSOR is used for marking both. Furthermore, adessive MANNER is seen in (58a) and REASON in (58b):

- (58) a. Rääki-si-n rõõmsa-l tooni-l  
 speak-PST-1SG cheerful-ADE tone-ADE  
 ‘I spoke in a happy tone’
- b. Sattu-si-n raske õnnetuse tagajärje-l haigla-sse  
 end.up-PST-1SG hard.GEN accident.GEN consequence-ADE hospital-ILL  
 ‘I ended up in a the hospital due to a bad accident’

Adessive REASON (58b) is distinguished from MANNER (58a) in that it answers the question *miks?* ‘why?’ as opposed to *kuidas?* ‘how?’. In addition, EVIDENTIAL refers to a small group of words, which often function as quotatives in reported speech, presenting the source of information (*tema sõnu-l* 3SG.GEN word.PL-ADE ‘according to him’). Adessive TIME includes *suve-l* summer-ADE ‘in the summer’. Adessive THEME includes the arguments of *avalduma* ‘be expressed in’ and *soovitama* ‘recommend’.

### Ablative functions

CONCRETE LOCATION includes (*tellingu-lt* scaffolding-ABL ‘off the scaffolding’ while ABSTRACT LOCATION refers to metaphorical Sources, e.g. *konto-lt* account-ABL ‘from the account’.

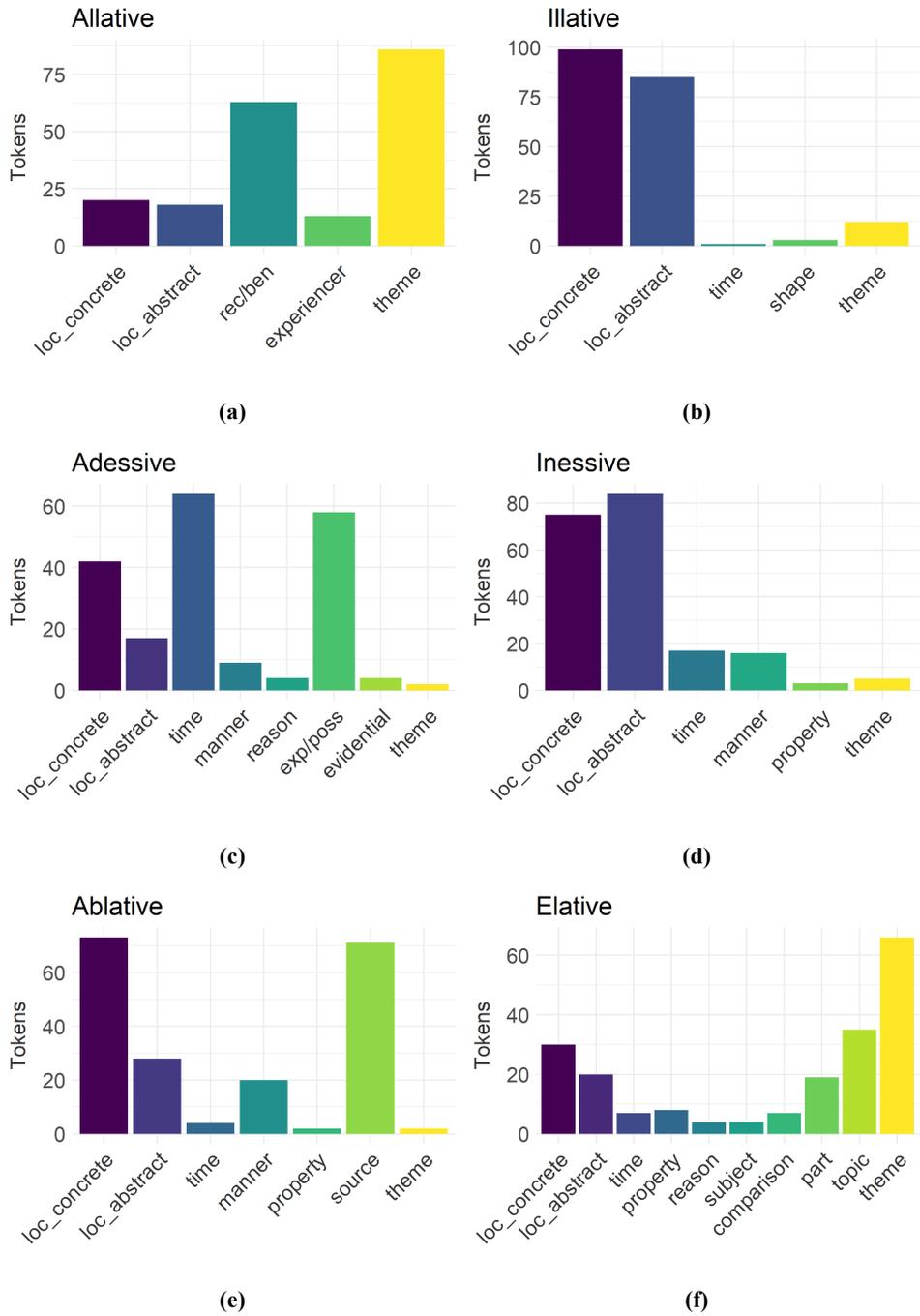
There is a certain degree of symmetry between the uses of allative and ablative. Namely, ablative expresses the role opposite to allative Recipients, i.e. human Sources, as shown in (59):

- (59) Kuul-si-n seda Klaudia-lt  
 hear-PST-1SG this.PAR Klaudia-ABL  
 ‘I heard about it from Klaudia’

Ablative MANNER phrases alternate with phrases marked by the postposition *poolest* ‘by’ (60a). Adessive PROPERTY is seen in (60b), TIME in (60c) and THEME in (60d):

- (60) a. Euroopa rikkuse-lt kolmas riik  
 Europe.GEN wealth-ABL third.NOM country.NOM  
 ‘Third country in Europe by wealth/Third richest country in Europe’
- b. Mees Niskamäe-lt  
 man.NOM Niskamäe-ABL  
 ‘A man from Niskamäe’
- c. Kursusekaaslane ülikooli päev-i-lt  
 course.mate.NOM university.GEN day-PL-ABL  
 ‘A course mate from university time’
- d. Ta taba-ti abielurikkumise-lt  
 1SG.NOM catch-IMPRS.PL adultery-ABL  
 ‘S/he was caught with adultery’

## 5.2.2 Results



**Figure 19:** Functional profiles of Estonian spatial cases. ‘exp/pos’ refers to the function EXPERIENCER/POSSESSOR and ‘rec/ben’ refers to the function RECIPIENT/BENEFICIARY.

The functional profiles of Estonian spatial cases are seen in Fig. 19. Cases were found to vary in terms of the number of different functions in which they are used, as well as the frequency with which their various functions occur.

One may note that the six graphs in Fig. 19 all include CONCRETE LOCATION and THEME. The former constitutes a highly semantically coherent role (with one particular type of meaning) while latter is semantically non-coherent, merely marking the participation of something in an event/state (i.e. the argument relation).

A vast majority of roles depicted in Fig. 19 may be regarded as either spatial or metaphorically spatial. Regarding spatial functions (CONCRETE SPACE and ABSTRACT SPACE, 'loc\_concrete' and 'loc\_abstract' in Fig. 19), they dominate in the usage of illative (92%, Fig. 19b) and inessive (79.5%, Fig. 19d), as well as describe most ablative tokens (50.5%). For the rest of the three cases, spatial functions make up less than half of their instances. Allative and elative mark spatial relations the least frequently (19% and 25%, respectively).

Most coded functions are metaphorically spatial. For instance, TIME can be interpreted as a metaphorical location on a temporal axis, PROPERTY (60b) marks something originating from a place or being made of a material. MANNER (58a), similarly to SHAPE (51b), describes an abstract pattern of doing something as a space in which one is located and REASON (53c) (similarly to the elative function SUBJECT) describes the causal force behind an event as a location, from which it metaphorically originates. EXPERIENCER/POSSESSOR constitutes the locations of emotions, while the function RECIPIENT/BENEFICIARIES represents the Goals of goods and interactions (the ablative SOURCE being its opposite). In addition, EVIDENTIAL describes the location from which presented evidence originates, while PARTITIVE (53b) expresses the set from which something originates as a location. Similarly to PARTITIVE, COMPARISON (53a) outlines the external context in which something is highlighted as a location (*keskmise-st suurem* average-ELA bigger.NOM 'bigger than the average').

Adessive occurs in metaphorically spatial function more frequently than other cases (70%), which marks TIME and EXPERIENCER/POSSESSOR in most of its instances. However, they also constitute a significant proportion of ablative (49%) and allative (38%).

Non-spatial functions include TOPIC (54) and THEME (e.g. the elative argument of *hoolima* 'care about'), the latter also being the only bleached (semantically non-coherent) function. TOPIC only occurs with elative, meaning that most of the instances of elative case occur with no spatial semantics (51%). However, THEME is the most dominant among elative functions, meaning that elative is entirely bleached in 33% of its instances. This is similar to allative, where it is also the most frequent function (43%). The other four cases occur in THEME function in a small minority of their instances (1%–8%).

Fig. 19 also shows that cases vary greatly in terms of the number of functions in which they occur. This ranges from five in the two lative cases (Fig. 19a–19b), to 10 for relative. It is interesting that it is the two lative (Goal-marking) cases that mark the least different functions. Inessive and ablative have 6 and 7 functions, respectively, while adessive has 8 and relative 10. Hence, relative is the most multifunctional case in Estonian.

### 5.2.3 Summary

Cases vary greatly both in terms of how many functions they have, as well as how frequently they mark bleached functions. Relative and allative stand out as the cases with the highest proportion of the bleached *THEME* function. Relative also stands out as the most multifunctional case. Both properties highlight them as most highly grammaticalised in a functional perspective.

This result is in line with what is known about the Estonian relative as it has been noted to be a highly multifunctional case (Erelt et al., 2007; Nurka, 2014; Viht and Habicht, 2019). Multifunctionality is cross-linguistically also expected of allative (Rice and Kabata, 2007), but in Estonian it is the least multifunctional spatial case.

Illative and inessive are functionally almost entirely spatial (*CONCRETE LOCATION* and *ABSTRACT LOCATION*). Ablative is divided between directly spatial and metaphorically spatial use while adessive is mostly used to mark metaphorical space. Allative is divided between metaphorically spatial and bleached functions, while relative is the least spatial case, marking direct and metaphorical space in a minority of its instances.

## 5.3 Study 2: Lexical selection patterns of cases

This study aims to answer RQ5: “Which spatial cases have the widest lexical selection patterns?” It describes the data, method and results of investigating the lexical selection patterns (or lexical variance) of spatial cases, both in their overall usage as well as only in oblique argument function. More bleached spatial case affixes with wider functionality are also expected to be used on a wider variety of lexemes. As explained in section 5.1.2, this study makes use of several measures of morphological productivity.

### 5.3.1 Data

This study is based on two datasets, because it investigates lexical variance in two different contexts — 1) all instances of a case and 2) oblique function of a case. While cases are expected to vary in terms of lexical variance when all their instances are considered, it is unclear if this variance also occurs in bleached function.

In addition to the six spatial cases, data from other cases is also presented for the first context, i.e. all case instances. Data is extracted from the Balanced Corpus of Estonian (see section 4.1.1). However, instead of simply including all case tokens, data is treated in a variable-corpus approach. This means that the analysis is based on an equal sample of affixes from each case to prevent overestimating the variance of low-frequency affixes (Gaeta and Ricca, 2006). Hence, lexical variance is summarised based on a 10,321 token sample for each of the 14 cases. This amount was chosen as this is the total number of tokens for abessive, the case with the smallest token count in that corpus. In addition, instead of summarising variance in a single random sample, it was calculated based on mean values of five different random samples per case (each containing 10,321 tokens), to obtain a more balanced view.

In order to look at lexical variance in oblique function, a separate set of samples of case-marked tokens was compiled. The set was based on the dataset used for the corpus study in chapter 4 (see section 4.1.3). Tokens marked in spatial cases were extracted from 500 sentences containing oblique argument verbs. An equal number of tokens for five spatial cases was randomly selected from each token list ( $n = 627$ ). These token samples were analysed for realised and potential lexical variance.

### 5.3.2 Method

Lexical variance is somewhat similar to morphological productivity in that it describes the extent to which a pattern (a case affix) is applied to variable parts (nominals). Hence, it can be assessed in an analogous way to morphological productivity measures, i.e. by simple type counts (realised productivity) and *hapax legomena* (potential and expanding productivity).

Realised lexical variance (as an analogue to realised productivity) refers to the extent of the current usage of a morphological pattern. It is measured by simple lemma type count ( $V$ ) (Baayen and Lieber, 1991; Baayen, 1993; Baayen and Renouf, 1996), reflecting the number of different lemmas with which the morpheme has been observed to occur in a particular corpus or dataset. This measure only describes the morpheme's lexical variance in the past, i.e. the time period covered by data.

Potential lexical variance (as an analogue to potential productivity) is based on *hapax legomena*. Hapaxes constitute forms which only occur once in the corpus. Hence, this measure integrates both type and token count to indicate availability to new contexts. Broadly, hapaxes reflect peripheral areas in the usage of a pattern, where it is not yet widespread, but might become more frequently used in the future.

Potential lexical variance expresses the proportion of a morpheme's hapaxes ( $nI$ ) among all tokens including that morpheme ( $Nc$ ). This means it shows the probability of a word marked in a particular morphological pattern occurring only

once in the corpus. If the P of one case category is higher than that of another, it indicates higher lexical variance compared to the other category.<sup>31</sup>

These two measures for assessing the lexical variance of case affixes are illustrated in Table 12.

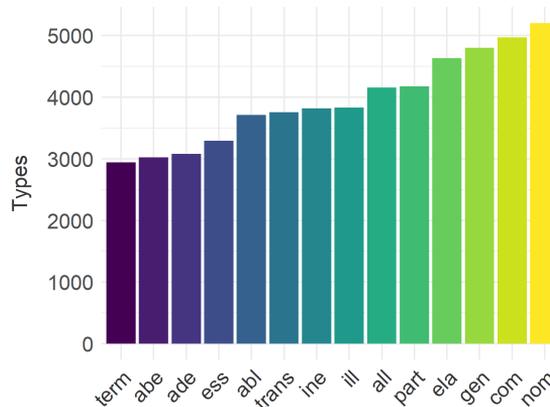
**Table 12:** Productivity measures used for assessing lexical variance.

Measure	Formula
Realised variance	lemma types ( $V$ )
Potential variance	$P = nI/Nc$

However, using potential variance in a variable-corpus approach means that it will be reflected by a simple hapax count. This is because all samples are equal in terms of token count ( $Nc$  in Table 12). Hence, for all affixes, hapaxes will be divided by the same number, effectively rendering hapax count as the indicator of potential variance.

### 5.3.3 Results: all case tokens

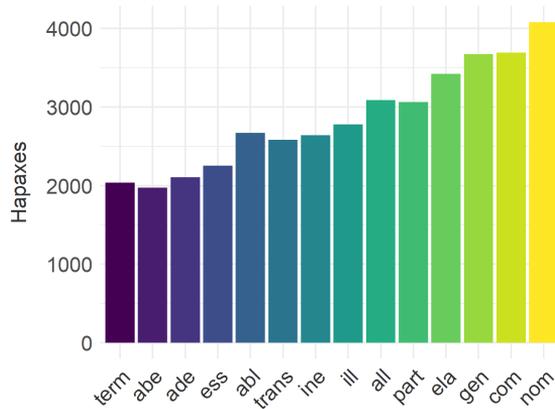
Fig. 20 shows the realised lexical variance of all Estonian cases in all functions:



**Figure 20:** Realised lexical variance of 14 Estonian cases. Figure presents mean values based on five random samples of 10,321 tokens for each case.

<sup>31</sup> Another hapax-based measure is expanding productivity, which reflects the proportion of a pattern's hapaxes ( $nI$ ) among all hapaxes in a corpus, reflecting the likelihood of a word belonging in a particular category, if it was found only once in the corpus. It is not used in the present study, because in a variable-corpus approach the number of all corpus hapaxes is unavailable.

As expected, syntactic cases have higher realised variance values than most semantic cases. However, elative, allative and comitative all have relatively high realised variance. Elative is significantly more lexically varied than partitive ( $\chi^2 = 41.178$ ,  $df = 1$ ,  $p < 0.0001$ ). Genitive and comitative are more varied than elative (genitive vs elative:  $\chi^2 = 5.5098$ ,  $df = 1$ ,  $p = 0.02$ ). There is no distinction between allative and partitive, but both are significantly more varied than the rest of the semantic cases (allative vs inessive:  $\chi^2 = 22.658$ ,  $df = 1$ ,  $p < 0.0001$ ). Overall type-token ratios for spatial cases range between 0.36–0.45.

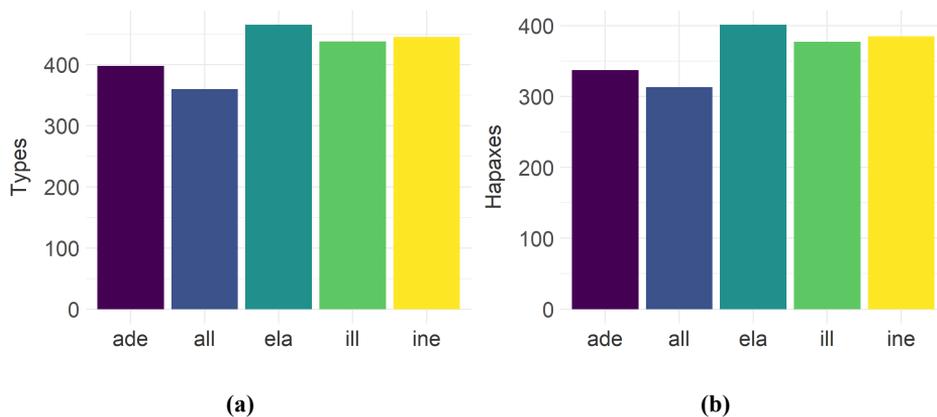


**Figure 21:** Potential lexical variance of 14 Estonian cases in the Balanced Corpus of Estonian. Figure presents mean values based on five random samples of 10,321 tokens, compiled for each case.

Fig. 21 presents a very similar picture for potential variance. Similarly to realised variance, elative has higher potential variance than partitive and allative (elative vs partitive:  $\chi^2 = 28.975$ ,  $df = 1$ ,  $p < 0.0001$ ) and lower variance than genitive and comitative (genitive vs elative:  $\chi^2 = 13.315$ ,  $df = 1$ ,  $p = 0.0003$ ). Illative, inessive and ablative do not differ from one another statistically in realised nor potential variance (Fig. 20–21), while adessive is the least varied spatial case in both measures.

### 5.3.4 Results: spatial cases in governed position

Compared to the overall view, the lexical variance of spatial cases is more uniform when only focusing on their structural function when they are selected by verbs:



**Figure 22:** Realised (a) and potential variance (b) of spatial cases in governed position.

Fig. 22a shows that in structural function, internal cases are significantly more lexically varied than external cases (illative vs adessive:  $\chi^2 = 5.1756$ ,  $df = 1$ ,  $p = 0.023$ ), but there are no statistically significant distinctions inside the internal or external case groups. Same pattern applies to potential variance (Fig. 22b) (illative vs adessive:  $\chi^2 = 4.9469$ ,  $df = 1$ ,  $p = 0.026$ ). Overall type-token ratios for spatial cases range between 0.57–0.75.

### 5.3.5 Summary

All in all, lexical variance in all case instances highlights relative and allative as the most lexically varied cases, while the oblique perspective highlights relative. This illustrates that the overall semantic coherence of these two cases is low, which is likely due to them occurring in a much wider range of contexts than other spatial cases. Relative even exceeds partitive in this measure, which is the most frequent object case in Estonian, as was shown in the corpus study (see Table 6 in section 4.1.4).

Oblique function was quantitatively shown to differ from overall case usage in terms of higher type/token ratios. For instance, relative has 0.45 types per 1 token in overall usage, but 0.75 types per 1 token in oblique function. Hence, it is indeed semantically less coherent, because all spatial cases are more lexically varied in oblique function than overall.

Allative, too, is more lexically varied in oblique function, but the distinction between its two datasets is smaller (0.40 types per token overall, 0.57 types per token in oblique function). This is likely due to the pronominalisation resulting

from its tendency to mark animate roles in oblique function. As was shown in Fig. 10, allative and adessive stand out among the spatial cases in terms of marking a higher proportion of human participants in oblique function. Lexemes with human animate referents are more likely to pronominalise than lexemes with inanimate referents, which likely reduces allative type-token ratio in oblique function, when compared to internal cases.

## 5.4 Study 3: Syntactic productivity

This section will answer RQ6: “Which oblique argument structures are the most productive?” It presents an overview of the syntactic productivity of Estonian argument structures by using a random sample of all Estonian verbs to investigate the proportion of different structures among them.

### 5.4.1 Data and method

All verb lemma types were extracted from the Balanced Corpus of Estonian ( $n = 7,651$ ). From the set of verbs with a relative frequency above 1 pmw ( $n = 3,043$ ), 1,000 random verbs were selected. Each verb was assigned an argument structure, describing 1) its number of arguments and 2) the cases marking these arguments. This means that the verb was mapped with a structure that specifies its most prominent event participants, which are partially defined by their semantic obligatoriness, partially by their fixed case (or case alternation system) or adposition. The sample is freely available (Aigro, 2022b).

While in most instances coding arguments was a rather straightforward process, issues occurred with some verbs. In order to offer a solution for them and to keep the coding as objective as possible, the following principles were followed:

- The single-participant structure coded as *nom* (nominative) was assigned to one-place verbs with a single nominative subject participant. This includes unergative verbs (*tantsima* ‘dance’) and unaccusative verbs (*sündima* ‘be born’/*tantsima* ‘dance’). The structure also entails various interaction verbs, e.g. *lõugama* ‘yell’. One borderline category within this set contains verbs making strong reference to concrete locations (*elutsema* ‘live’/‘habitate’, *kaugenema* ‘distance from’) or manner (*käituma* ‘behave’). As most of them do not occur with a specific case (for instance, *elutsema* ‘live’/‘habitate’ may occur with both inessive and adessive phrases), they were included under the *nom* structure.
- Structure *nom\_can* was assigned to verbs with a nominative subject and a canonically marked object, e.g. *mõistma* ‘understand’, *solvama* ‘insult’ and *loopima* ‘throw’. Partitive verbs (those not allowing DOM) are not distinguished from other canonically marking verbs.

- Verbs with a mandatory adposition were coded as *nom\_adp*, e.g. *klammerduma* + *külge* ‘cling’ + ‘onto’ and *protesteerima* + *vastu* ‘protest’ + ‘against’.
- Three arguments (*nom\_can\_all*, *nom\_can\_el*, etc) were coded when the meaning of the verb requires three participants. This includes quintessential indirect object verbs (e.g. transaction verbs such as *laenama* ‘borrow’, *annetama* ‘donate’), and Goal or Source focused verbs with abstract Goals and Sources (*ammutama* ‘draw’/‘extract’).
- Verbs with oblique arguments were coded as selecting structures *nom\_ad*, *nom\_el*, *nom\_com*, etc, depending on the oblique case. If a verb was also found to be included in the corpus-based dataset in chapter 4 (see section 4.1.3), it was coded as selecting the case with which it was coded in chapter 4.
- Verbs only selecting infinitive forms of verbs as arguments were coded as *nom\_verb*, e.g. *vaevuma* ‘bother [to do something]’ and *võima* ‘be allowed’/‘can’.
- Verbs with a single non-nominative participant marked in a structural case, e.g. weather verbs (*sadama* ‘rain’/‘snow’), were coded as *can* (= a canonical object case).

In addition, the study did not differentiate between word order conventions when coding structures. Hence, all structures including nominative mark it in the first position (*nom\_can*). Hence, *meeldima* ‘like’ and *osutama* ‘point at’ are both coded as having the same structure (*nom\_all*), even though the Experiencer of *meeldima* ‘like’ is more likely to occur preverbally than the Goal of *osutama* ‘point at’.

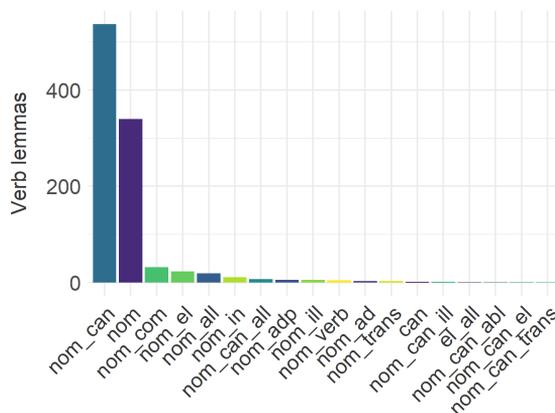
## 5.4.2 Results: syntactic productivity in general

Overall, 18 different argument structures were found in the sample of 1000 random Estonian verb lemmas (described in section 5.4.1). Their proportions are presented in Table 13:

**Table 13:** Argument structures among 1000 random Estonian verbs (< 1pmw).

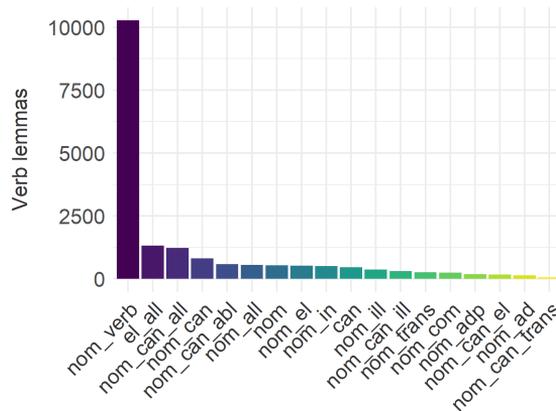
Argument structure	Proportion	Example
nom_can	53.7%	<i>sulatama</i> ‘melt’
nom	34%	<i>magama</i> ‘sleep’
nom_com	3.2%	<i>piirnema</i> ‘border with’
nom_el	2.3%	<i>eristuma</i> ‘differ from’
nom_all	1.9%	<i>halastama</i> ‘have mercy on’
nom_in	1.1%	<i>pettuma</i> ‘become disappointed in’
nom_can_all	0.7%	<i>annetama</i> ‘donate’
nom_adp	0.6%	<i>hääletama</i> ‘vote’
nom_ill	0.6%	<i>sekkuma</i> ‘become involved with’
nom_verb	0.5%	<i>vaevuma</i> ‘bother’
nom_ad	0.3%	<i>kripeldama</i> ‘bug’/‘annoy’
nom_trans	0.3%	<i>valmistuma</i> ‘get ready for’
can	0.2%	<i>sadama</i> ‘rain’/‘snow’
nom_can_ill	0.2%	<i>kaasama</i> ‘involve sb in’
el_all	0.1%	<i>piisama</i> ‘suffice’
nom_can_abl	0.1%	<i>laenama</i> ‘borrow’
nom_can_el	0.1%	<i>ammutama</i> ‘extract’/‘draw’
nom_can_trans	0.1%	<i>tituleerima</i> ‘award sb the title of’

Fig. 23 illustrates the proportions of argument structures in the verb sample:



**Figure 23:** Argument structure productivity in a sample of Estonian verbs (n = 1000).

The two most canonical structures, *nom* for verbs expressing events or states with one participant and *nom\_can* for those with two participants, constitute the most frequent structures. They are used by 34% and 54% of Estonian verbs, respectively. Other than these two, oblique frames with comitative (3.2%), elative (2.3%) or allative (1.9%) arguments are the most predominant oblique structures. There is no statistically significant distinction between their frequencies in the sample. The rest of the spatial case frames are much less frequent (0.3–1.1%). Similarly to the study in chapter 4, ablative was not found in the structure of any one-place verb or two-place verbs.



**Figure 24:** Mean verb frequency per argument structure in Estonian verbs (n = 1000).

Fig. 24 outlines mean verb frequency per each structure found in the sample. Essentially, these numbers represent structure token counts. The *nom\_verb* structure has the highest mean token due to the modal verb *võima* ('may'/'can'). It is followed by a mixture of canonical and oblique structures.

### High- and low-level extensions

As discussed in section 5.1.3, there are several ways in which argument structures can have the potential to be extended to new verbs (Barðdal, 2008). High-level extensions refer to semantically bleached uses of case, involved in patterns which exhibit high type count. Such structures are assumed to be available for new verbs (i.e. productive), because their overall meaning is highly bleached, as is indicated by their higher type frequency (number of different verbs).

Low-level extensions, on the other hand, refer to verb groups with low type counts. Such verbs are expected to be highly coherent in terms of meaning. Regardless of low type count, such structures may also be extended to new verbs if the novel verb exhibits semantic similarities to the verbs already used with the

structure. In such instances, verb groups with higher average token frequencies are expected to be more available for extension than verb groups with lower token frequencies.

Section 4.2.2 showed that this might be the case with verbs in inessive and adessive structures, which were found to be highly semantically coherent. In the oblique argument dataset, nine out of ten inessive verbs and four out of five adessive verbs were shown to be near-synonyms.

The two sets are also similar to one another. Both refer to the abstract representation of something in something else (e.g. *peegeldu-b seaduste-s* be.reflected-PRS.3SG law.PL-INE ‘is reflected in laws’, *baseerub seadustel* be.based-PRS.3SG law.PL-ADE ‘is based on laws’). Hence, the two patterns may be in competition as potential analogy-based extensions for the same novel verbs.

Both of these structures had relatively low type counts in a random corpus sample of 1000 verbs (11 for inessive and 3 for adessive), which is characteristic to low-level extensions. However, as seen in Fig. 24, inessive-selecting verbs are more frequent on average (mean verb frequency = 516.1 tokens in the corpus) than adessive-selecting verbs (mean verb frequency = 147.7). This means that the study predicts further extension for the inessive argument structure and less so for the adessive structure.

Illative verbs, on the other hand, constitute a rather varied group, including several smaller semantically coherent groups. Most of them convey metaphorical spatial semantics. This includes verbs describing becoming involved with something (*sekkuma* ‘become involved with’, *sukelduma* ‘dive into’), or with another person (*armuma* ‘fall in love with’), as well as actions with abstract Goals (*investeerima* ‘invest in’, *kaasama* ‘involve in’). As a low-level extension it might be extended to novel verbs with clear and present spatial Goal meaning.

Finally, allative and elative argument structures have high-level extension potential. The verbs occurring with them are highly semantically varied in terms of the participant roles they encode. These structures also have the highest verb type counts among all spatial case structures. When considering the results from chapter 4, it is likely that these verbs are available to K-states (property and relationship verbs) as well as other non-dynamic events, especially those involving more than one human participant. All these features were more prominent in the oblique dataset than among verbs selecting canonical objects.

### 5.4.3 Summary

This section investigated the syntactic productivity of argument structures in Estonian. At least 18 different argument structures are used in Estonian, with the top two (canonical) structures occurring with 88% of verbs. This means that 12% of Estonian verbs use some type of non-canonical case marking on their arguments.

The three predominant oblique structures include comitative, elative and allative complements. The data presented here, together with the results of the corpus

study in chapter 4, the function study in section 5.2 and the variance study in section 5.3 give reason to regard elative and allative as already showing some properties of high-level extensions. Even though the productivity of their structures is much lower than that of the canonical object structure, they nevertheless occur with a variety of different semantically non-coherent verbs and mark various roles.

The rest of the structures with spatial cases might constitute low-level extensions. Given the semantic overlap between verbs selecting inessive and adessive arguments, these two structures are likely to be in competition with one another in which regards novel verbs that are semantically similar to verbs already selecting inessive and adessive. The results of this study predict that inessive structure is more available to novel verbs than adessive structure.

## 5.5 Summary

The main aim of the three studies presented in this chapter was to assess the degrees of grammaticalisation of the six Estonian spatial cases. The results of all three studies highlight elative and allative as the most grammaticalised spatial cases. This complements the results of chapter 4, which showed that verbs selecting elative and allative are most like verbs selecting canonical objects, both in terms of various lexical semantic properties as well as usage patterns.

Their higher degree of grammaticalisation was concluded from the three studies presented in the present chapter. First, elative and allative occur in semantically bleached functions more frequently than other spatial cases. Both lack any, even metaphorical spatial meaning in approximately half of their overall instances, elative in more than half. In addition, elative is the most multifunctional spatial case with the most varied profile.

Second, elative and allative occur with the widest range of nominals. Elative even exceeds the canonical object case, partitive, in this measure, while allative variance is similarly to that of partitive. Third, elative and allative argument structures constitute the most productive syntactic argument structures among spatial cases. The only oblique argument structure more productive than elative and allative is a structure where arguments are marked in comitative. The reasons for this are discussed in section 5.5.2.

A higher degree of grammaticalisation of a case not only indicates higher productivity for its argument structure, but it also affects the way the structure is used. Section 4.4 compared the usage of oblique argument structures to that of the canonical structure with direct objects. It found that the elative structure is used in the most similar way to the canonical structure in terms of a variety of factors, e.g. argument nominal number, verb tense proportions and NP length in lexemes. Hence, case functionality (grammaticalisation degree) is interlinked with how often it occurs in structural position, as well as *how* it is used in structural position.

These results are also interesting from the perspective of allative functions.

The oblique use of allative is often regarded as overlapping with the functions of dative in Indo-European languages, marking human Goals of transaction verbs and Experiencers of emotion verbs (Lindström and Vihman, 2017). However, the functional profile of allative (Fig. 19a) shows that it occurs in the entirely bleached THEME function more often than in RECIPIENT/BENEFICIARY and EXPERIENCER put together. In addition, the structure *nom\_all* was more predominant among Estonian argument structures than a structure where allative marks the third participant (*nom\_can\_all*). Hence, although the Recipient and Experiencer roles stand out as they refer to semantically specific entities, the Theme role is actually more frequent with allative, meaning it is a highly bleached case (rather than a case with dative-like function), marking event participation in almost half of its instances. This is a result that only a quantitative approach is able to find, because example-based approaches may be biased towards overemphasising semantically coherent roles and underestimating the proportion of entirely bleached instances.

### 5.5.1 From spatial meaning to structural use

Based on the three variables studied in this chapter — i.e. functional profiles, lexical selection patterns and syntactic productivity — one may postulate a pathway by which a semantically coherent spatial case becomes a bleached event/state participant marker. Table 14 describes the six phases on this pathway:

**Table 14:** Pathway leading from a spatial affix (phase 1) to a grammatical marker (phases 5–6)

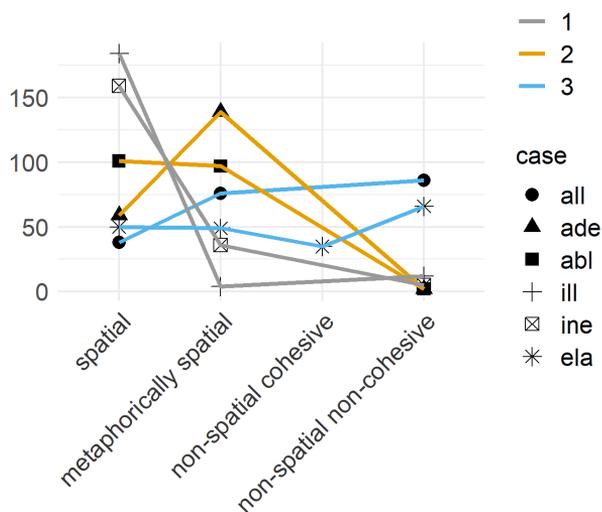
	Affix functions	Types of nominals	Verb examples
1	CONCRETE LOCATION	Concrete inanimate nominals ( <i>laua-l</i> table-ADE ‘on the table’), proper names of places ( <i>Belgia-s</i> Belgium-INE ‘in Belgium’)	<i>patrulli-b tänava-l</i> patrol-PRS.3SG street-ADE ‘S/he patrols on the street’
2	ABSTRACT LOCATION	Abstract inanimate nominals ( <i>mõtte-s</i> thought-INE ‘on one’s mind’)	<i>investeeri-b ettevõtte-sse</i> invest-PRS.3SG company-ILL ‘S/he invests in a company’
3	Semantically coherent, metaphorically spatial functions, e.g. TIME, PROPERTY, MANNER and PARTITIVE	Inanimate abstract nominals ( <i>laupäeva-l</i> Saturday-ADE ‘on Saturday’), animate nominals ( <i>helista ema-le</i> call.IMP.SG mum-ALL ‘call mum!’)	Verbs selecting Experiencers, Possessors and Recipients (e.g. <i>viirastuma</i> ‘imagine’, <i>laenama</i> ‘lend’)
4	Semantically coherent non-spatial functions (e.g. TOPIC)	All types of nominals	Verbs of interaction ( <i>räägt-n koera-st</i> talk-PRS.1SG dog-ELA ‘I am talking about a dog’)
5	THEME: semantically non-coherent, unaffected by the event or state described by the verb PATIENT: semantically non-coherent, affected by the event or state described by the verb	All types of nominals	<i>sõltuma</i> ‘depend on’, <i>mõistma</i> ‘understand’
6		Concrete animate and concrete inanimate nominals	<i>parandama</i> ‘mend’

The pathway outlined in Table 14 uses functions, congruent nominal types and verbs to describe the phases a spatial case may pass through on its way to a bleached grammatical marker. The original meaning of spatial cases is represented by phase 1, while phase 6 signifies the furthest point from it.

When moving from phase 1 to 2 and from phase 2 to 3, a case begins occurring with non-congruent nominals, which for a spatial case include abstract inanimate nominals and human nominals (Aristar, 1997). However, when the case gains the even more bleached functions of phases 4, 5 and 6, the nominal types generalise, meaning that congruent nominals (inanimate concrete) again occur with the case affix in these functions. This is possible, because the spatial case affix has become bleached enough for it to not automatically trigger spatial interpretation when occurring with inanimate concrete nominals. Instead, structure meaning overrides the meaning merely derived from the combination of the noun and the affix (phases 4, 5 and 6 all describe morphological selection, e.g. by verbs). Hence, when speakers encounter the sentence *sõltu-b hoone-st* depend-PRS.3SG building-ELA ‘depends on the building’, they do not first interpret the word *hoone-st* (building-ELA) to refer to something moving out of the building, but rather interpret it only in the context of ‘depend on’.

As is evident from Table 14, congruent nominals become limited again in phase 6, where most prototypically transitive events must include concrete nominals, for them to maximally affect the event and be maximally affected by it. Hence, it might be problematic to view phase 6 as more grammaticalised than phase 5. On the one hand, canonical object cases in Estonian were shown to have the functionality of phase 6, while oblique cases were not found to mark Patients in section 4.3 (with the allative argument of *virutama* ‘hit’ being the closest candidate). However, even among canonical object verbs, Patient role is much more rare than the Theme role. In addition, it entails more semantics than merely participating in the event/state described by the verb.

Based on Table 14 and the functional profiles of Estonian cases outlined in Fig. 19, this thesis is able to describe their progression along these six phases. This progression is shown in Fig. 25:



**Figure 25:** Estonian spatial cases on a grammaticalisation pathway. The six phases of the pathway are collapsed into four levels on the x axis of Fig. 25. Phases 1 and 2 in Table 14 are both in the ‘spatial’ category. Phase 3 is represented by ‘metaphorically spatial’, phase 4 by ‘non-spatial cohesive’ and phases 5–6 are found in the ‘non-spatial non-cohesive’ group. Groups 1–3 represent cases behaving in a similar fashion and having broadly the same usage pattern.

Based on the functional profiles, lexical variance and syntactic productivity of Estonian spatial cases, illative and inessive (type 1 in Fig. 25) cases are still mostly in the early stages of the pathway outlined in Table 14. They mostly mark space, either in a concrete or abstract way (both *CONCRETE LOCATION* and *ABSTRACT LOCATION* belong under the ‘spatial’ category in Fig. 25).

Ablative and adessive (type 2 in Fig. 25) are somewhere in the middle stages. The somewhat higher degree of grammaticalisation for adessive is evident from two aspects of its use: 1) adessive occurs as part of the argument structure of two-place verbs, while ablative was not found in such a structure (Table 13) and 2) spatial functions are less frequent for adessive than for ablative. This means adessive has taken on more frequent metaphorically spatial functions, such as *EXPERIENCER/POSSESSOR*.

Finally, elative and allative (type 3 in Fig. 25) have reached the highest phases of the grammaticalisation pathway. Spatial function makes up a quarter of their instances or less. In about half of their instances, spatial function is not present, not even metaphorically.

Similarly to the study by Nichols and Timberlake (1991), the studies in this chapter show the complex nature of grammaticalisation. First, a case need not develop all of the functions described in Table 14, or it may do so marginally (e.g. Estonian inessive is only infrequently used as an argument marker).

Second, when a case does develop further along the phases described in Table 14, it does not simply pass through them as is suggested by the metaphor of a grammaticalisation *pathway*. Instead, the process is described by the broadening of functionality. This means that a case need not leave the earlier phases behind when it becomes more common in bleached phases. Instead, it may become a highly multifunctional affix that works well in a wide range of domains. It resists the semantic and structural case labels as it functions as both. Hence, the description of a case system might benefit from an approach where one initially examines cases in terms of the proportions of potentially monofunctional and multifunctional cases, rather than the proportions of semantic and structural cases. This, in turn, would help postulate an empirically more sound argument structure system, rather than being concerned with argument marking being the sole function of only a few cases.

The complexity of grammaticalisation is also evident from the fact that the meaning of cases may become extended via distinct routes, even inside the framework in Table 14. For instance, phase 3 has allowed the meaning of Estonian spatial cases to become generalised in different ways. Allative and adessive were extended towards metaphorically spatial human participants, while elative meaning was extended towards metaphorically spatial non-human functions (e.g. PARTITIVE, COMPARISON). Hence, roles entailed in these phases vary, leading to cases that are rather different in terms of specific functions (semantic roles), but similar in terms of their more general profiles (the phases of Table 14).

### 5.5.2 From comitative meaning to structural use

In addition to spatial cases, comitative case was found to figure in a relatively productive oblique argument structure. Among 1000 random Estonian verbs, 32 verbs were found to select comitative arguments (marked in *-ga*). The group is large enough to provide us with some understanding of the different ways in which comitative meaning can be represented in the meanings of verbs.

Comitative is inherently different from spatial cases in that its type-congruent nominals refer to animate entities. This is because its most canonical function is to mark plurality on the part of human event/states-of-affairs participants. In this function, comitative indicates a somewhat equal status of event participants and it alternates with plural morphology (*ema ja mina* mum.NOM and 1SG.NOM ‘mum and I’ vs *ema-ga* mum-COM ‘with mum’).

**Table 15:** Verbs selecting comitative arguments.

Verb (Estonian)	Verb (English)
amelema	'make out with'
assotsieeruma	'associate with'
kõnelema	'talk with'
samastuma	'relate to'/'identify with'
piirnema	'border with'
seksima	'have sex with'
seostuma	'relate to'
suhetama	'relate to'
manipuleerima	'manipulate with'
võitlema	'fight with'
sõnelema	'have words with'
kattuma	'become covered with'/'be the same as'
uhkustama	'boast'
äritsema	'haggle'/'sell'
ristuma	'cross'
semmima	'date'
kättelema	'shake hands with'
rahulduma	'be content with'
sõbrunema	'become friends with'
kurameerima	'date'
vehkima	'wave around'
riidlema	'quarrel with'/'tell off'
liialdama	'exaggerate'/'overdo'
seonduma	'relate to'/'be connected with'
armatsema	'make love to'
leppima	'make up with'/'settle with'
vestlema	'converse with'/'talk to'
ühtima	'be the same as'
mehkeldama	'curry favour with'
lõikuma	'intersect with'
tutvuma	'become acquainted with'
lävima	'communicate with'

Table 15 presents the comitative-selecting verbs in the study. Some of these verbs appear to select comitative in a function where it marks mere plurality, referring to several (human) participants, all of whom participate in the event in a rather equal manner. Such verbs include *kurameerima* 'date', *lävima* 'communicate with someone' and *tutvuma* 'meet'/'become acquainted with'. In these events, (human) participants are involved in an equal manner, as is reflected in the fact that in these contexts the comitative structure may be substituted with a plural or a conjunction phrase. Hence, the three examples in (61a)–(61c) all have the same meaning:

- (61) a. Anna kurameeri-b Mari-ga  
 Anna.NOM date-PRS.3SG Mari-COM  
 ‘Anna is dating Mari’
- b. Anna ja Mari kurameeri-vad  
 Anna.NOM and Mari.NOM date-PRS.3PL  
 ‘Anna and Mari are dating’
- c. Nad kurameeri-vad  
 3PL.NOM date-PRS.3PL  
 ‘They are dating’

The extension of comitative begins with contexts where one of the two human participants is more agentive. In such instances, comitative is semantically more bleached than in (61). This is the case with *riidlema* ‘tell someone off’/‘scold’ (vs the more equal setting of *sõnelema* ‘quarrel with’/‘argue with’).

Comitative meaning is even further bleached in verbs describing events which generally only involve one human participant, the other participant being inanimate (e.g. *vehkima* ‘wave around’). However, some degree of comitative meaning is present with an inanimate concrete participant, because even though only one participant is animate, both still physically occur together in the same setting. However, when the inanimate argument is abstract, comitative meaning can be said to be lost, e.g. *uhkustama* ‘boast’ in (62):

- (62) Ma uhkusta-sin oma kraadi-ga  
 1SG.NOM boast-PST.1SG REFL degree-COM  
 ‘I boasted my degree’

In any case, in the most semantically coherent end of the spectrum, comitative is selected by the verb to mark the plurality of participants (61). In the other end of the spectrum, verbs select comitative in bleached function, describing a relationship between two participants, one of which may be animate (62). Hence, comitative starts out in an animate domain, but over time becomes bleached enough to also occur in an inanimate domain. The reverse is true for spatial cases, which start out in inanimate space, being then also extended into animate space. This change in animacy patterns affects some more than others, but as was seen in section 4.4.3, even elative which marks no explicitly animate roles, marks humans in 13% of the time when occurring in oblique function.

## 6 ARGUMENT STATUS ON A GRADIENT

This chapter uses behavioural data to investigate the argument status of oblique arguments and several other types of constituents marked in spatial cases. Up to this point, this thesis has referred to verbal dependents in oblique structures as ‘arguments’ (e.g. the argument of *sõltuma* ‘depend on’) following the tradition of viewing them as semantic arguments (Erelt, 2017c; Ackema, 2015). However, various approaches are divided in terms of the argument status of oblique arguments: some assume it is qualitatively different (weaker) from that of core arguments (Blake, 2001; Erelt, 2004; Barbu and Toivonen, 2016). Others expect both to express equally strong argumenthood (Nichols, 1983). Hence, up to this point the present thesis has used the term ‘argument’ to refer to both oblique arguments and canonical objects, but has made no explicit claim that the term has the same meaning for both on all levels. This chapter, however, will address exactly this question: do structures differ in terms of argument status quality?

Experimental approaches have successfully used various measures for assessing argument status. Verbal argument structure has been noted to play an important role in sentence processing as experimental paradigms have highlighted various distinctions between quintessential arguments and adjuncts. Such approaches have shown that a deep cognitive link exists between verbs depicting events/states and nominals expressing event/state participants — a link that is absent between verbs and quintessential adjuncts (further discussed in section 6.1.1) However, these measures have not been used for comparing oblique and canonical argument structures, i.e. contexts where verbs select a particular case for marking their arguments.

The results presented in the earlier chapters of this thesis (chapters 4–5) do not exclusively support either hypothesis on the argument status of oblique arguments. On the one hand they suggest that both canonical and oblique structures mark arguments of the same type/strength. This is because they showed that Estonian spatial cases, especially elative and allative, have extensive bleached function, marking the participants of a number of verbs without any trace of spatial semantics. They are the most grammaticalised spatial cases and even the usage patterns of elative-selecting verbs are similar to those of canonical object verbs. One might therefore expect the argument status of oblique arguments (or at least that of elative or allative arguments) to pattern similarly to that of canonical arguments.

On the other hand, chapter 4 also showed that verbs occurring in oblique argument structures are generally less transitive than verbs occurring with canonical objects. They are much more likely to refer to states expressing properties or relationships (K-states), but they are also more likely to refer to non-dynamic events. In addition, chapter 5 showed that even though elative and allative are more grammaticalised than other spatial cases, argument structures incorporating these cases are much less productive than canonical structures. These notions suggest an ar-

gument status distinction between oblique and canonical structures.

By incorporating various types of constituents in spatial cases, this chapter aims to investigate this question. It hypothesises that argument status is an inherently gradient phenomenon. This is confirmed by showing how particular types of semantically coherent categories occupy different parts of the argument status gradient, leading from non-argument constituents with coherent and concrete spatial semantics to arguments with non-coherent, bleached semantics. In this gradient view, the present study expects oblique arguments to pattern as strong arguments, but it also expects other types of (more spatial) elements to exhibit weaker argument status.

This chapter will first introduce the background for investigating argument status by experimental means (section 6.1) and then outline the aims and research questions of the experiments presented in this chapter (section 6.2). The design of the study is presented in section 6.3 and the results in 6.4. Finally, results are discussed in section 6.5.

## **6.1 Arguments in experimental approaches**

Instead of relying entirely on the morphosyntactic properties of argument structures to determine argument status, experimental approaches often make use of some other variable to indicate argument status, such as reaction times, eye movements or various offline measures. Section 6.1.1 will outline how such measures have been helpful in outlining the role of lexical argument structure in sentence processing. Section 6.1.2 will introduce the online and offline measures used for measuring argument status (section 6.1.2).

### **6.1.1 Sentence processing**

Argument structure, i.e. a linguistic device linking event markers (verbs) to participant markers (arguments) has been shown to play a fundamental role in sentence processing (see an overview in Tutunjian and Boland, 2008). A number of studies have indicated that during processing, verbs introduce information about potential event participants into the clause. This means that argument structure information is accessed as soon as the verb is recognised (Maurer and Koenig, 1999; Maurer et al., 2002). After encountering the verb, event participants are, in a way, expected to occur in the sentence as well. This is evident from the fact that speakers process postverbal argument phrases more quickly than postverbal adjunct phrases (Clifton et al., 1991; Speer and Clifton, 1998; Kennison, 1999; Schütze and Gibson, 1999). Furthermore, Kennison (2002) showed that when one-place verbs (e.g. ‘perform’) occur with an object they are not expected to encode in their argument structure (‘perform a play’), these objects pattern similarly with adjuncts. This type of difference in processing arguments and adjuncts has also been demonstrated in German (Konieczny et al., 1997) and Japanese (Yamashita, 1995).

It has been subject to debate, whether both arguments and adjuncts are attached via the same mechanism (MacDonald et al., 1994), or whether the two differ in their attachment mechanisms, arguments being facilitated by lexical knowledge and adjuncts by non-lexical real-world knowledge (Boland and Boehm-Jernigan, 1998; Boland, 2005; Tutunjian and Boland, 2008). The first hypothesis posits that the observed difference in the processing speeds of arguments and adjuncts does not result from some type of syntactic/lexical status of arguments, i.e. the fact that the former are included in the lexical information pertaining to verbs and the latter is not. Instead, it derives from co-occurrence frequency and typicality. In this view, arguments are not essentially different from adjuncts, because frequencies and typicality facilitate the attachment of both. Hence, the quicker processing of arguments is a result of the fact that arguments are common co-occurring elements of verbs than phrases expressing other types of information, e.g. Beneficiaries or Locations.

In the other view, however, the observed effects are due to the reality of argument structure rather than frequency. It regards verbal arguments as attached via a lexical operation (i.e. argument structures encoded in the verb's lexical entry) while adjuncts are attached via a distinct cognitive operation. This is because adjuncts are not expected to be specified in the verb's argument structure. The differences in their processing speeds results from the fundamental distinction between these two operations.

Both the frequency-based hypothesis (Spivey-Knowlton and Sedivy, 1995; Knoeferle et al., 2008) and the argument structure hypothesis (Boland and Boehm-Jernigan, 1998; Traxler and Tooley, 2007) have received support in experimental evidence. The argument structure of verbs assigns certain other constituents a prominent position (Kennison, 2002), but participants have also been shown to use real-world (contextual) knowledge, which can override lexical argument structure (Knoeferle et al., 2008). This has led some theories to conjoin the two effects (Altmann and Steedman, 1988; Tutunjian and Boland, 2008). They suggest that both contextual effects and argument structure effects arise simultaneously during processing. As summarised by Tutunjian and Boland (2008, 638): '[...] grammatical knowledge proposes, real-world knowledge disposes [...]', meaning that argument structure knowledge highlights potential arguments while situational knowledge guide the selection of the most contextually likely candidate.

Tutunjian and Boland (2008, 642) name a number of relevant problems for future studies investigating argument structure, one of which is: "Is the argument/adjunct distinction binary? Is it categorical?" Answering them would contribute information that is "central to understanding how syntactic and semantic knowledge is represented in the mind and accessed during sentence comprehension." (Tutunjian and Boland, 2008, 642).

## 6.1.2 Measuring argument status

Most online studies on argument structure use various types of reading experiments. Online methods include self-paced reading and eye-tracking (Clifton et al., 1991; Schütze and Gibson, 1999; Koenig et al., 2003; Conklin et al., 2004; Grodner et al., 2005), which offer ways for quantifying the speed of processing. Others investigate arguments by means of a visual world paradigm. For instance, Boland (2005) presented participants with pictures while they listened to sentences, tracking their eye movements across pictures. It allowed her to investigate whether hearing a verb leads the participant's gaze onto pictures depicting potential arguments.

However, online methods entail strict restrictions, which means they may not be suitable for comparing all structures. This is discussed in the following section that will briefly discuss an attempt made for the present thesis to use self-paced reading for investigating the argument status of oblique arguments.

### **A non-successful experiment: using online measures for investigating oblique arguments**

This section describes the problems which occurred with a study initially designed for the present thesis. The study attempted to compare oblique and canonical arguments in a self-paced reading paradigm. These issues are relevant to argument structure studies and highlight offline methods as a more helpful means of investigating the phenomenon.

In the study, experimental items included sets of almost identical three sentences. In each set, the first sentence contains a verb with a direct object (63a). In the second sentence, the verb selects an oblique argument (63b). In the third sentence, the verb is presented with a canonical adverbial (63c). One such triplet is presented in (63):

- (63) a. MTÜ      pett-is              nõukogu    eelmise-l aasta-l  
          NGO.NOM deceive-PST.3SG council.PAR last-ALL year-ALL  
          ‘The NGO deceived the Council last year’
- b. MTÜ      allu-s              nõukogu-le eelmise-l aasta-l  
          NGO.NOM obey-PST.3SG council-ALL last-ALL year-ALL  
          ‘The NGO obeyed the Council last year’
- c. MTÜ      sära-s              nõukogu-s eelmise-l aasta-l  
          NGO.NOM shine-PST.3SG council-INE last-ALL year-ALL  
          ‘The NGO shined in the Council last year’

26 such triplets were created. Each experimental item was made up of five words.<sup>32</sup> Oblique argument verbs (63b) were selected from the corpus study described in chapter 2. Two-place verbs with canonical objects (63a) and one-place verbs (63c) were selected to match the verb lemma frequency of oblique argument verbs.<sup>33</sup> The only variable parts in each triplet were 1) verb lemma and 2) case marking on the following nominal.

The participants first saw a sentence where all words had been substituted with underscores (one per letter). By clicking a mouse button, the first word of the sentence was revealed. By clicking the button again, the next word was revealed while the first word was again substituted with underscores (the ‘moving window’). This provided data on how many milliseconds the participant spend processing each of the five segments. In order to ensure that participants focused on what they read, a yes/no question was asked about the contents of the sentence, occurring once every five sentences.

The main hypothesis was that upon encountering the verb in the second segment, the reading time of the third word would be quicker for the canonical object (63a) than for the canonical adverbial (63c) (although this distinction may also show up in the reading times of the 4th and 5th segments). Oblique arguments (63b) were predicted to be distinct from adverbials, either being processed as fast as canonical objects or mapping somewhere between these two categories.

The main issue became evident already before the launch of the experiment — namely, experimental items included sentences with questionable acceptability. All experimental items, including the triplet in (63), were tested with an acceptability judgment task using a 7 point Likert scale. Some were broadly judged as non-acceptable. For instance, sentences with a mean score under 4 included five items from among canonical object verbs, five items from oblique argument set and seven items from the set where verbs occurred with adverbials. Furthermore, very few sentences (7 out of 78) had a mean acceptability judgment above 6.0, including only three oblique argument items, two canonical object items and two adverbial items. All in all, while most experimental items were judged acceptable (scoring above 4), they were nevertheless regarded as somewhat odd.

In terms of results, oblique arguments exhibited the longest reading times, both in the 3rd and 4th word segments. However, stimuli generally varied a lot and showed rather inconsistent patterns. This is likely due to the non-sufficient acceptability of experimental items. It is questionable whether it is possible to

---

<sup>32</sup> This design describes the second attempt at this experiment. The first design, piloted with 10 participants, included sentences that were made up of four words. It was altered to increase the number of segments on which the distinction in reading time may become expressed, as well as to alter stimuli so that they would be more natural.

<sup>33</sup> Based on a 50 sentence sample for each verb from the Balanced Corpus of Estonian, it was confirmed that all two-place verbs in the sample occur with an overt object in more than half of their instances. Similarly, one-place verbs were confirmed to occur without them in more than half of their instances.

design stimuli restricted in this manner to have high acceptability ratings, because the distinct transitivity patterns of these verb types mean they are difficult to fit in the same sentence context. For instance, as shown in chapter 4, oblique verbs often constitute states, which do not combine well with time and location adverbials, rendering odd interpretations.

Hence, this experiment is regarded as not having produced a negative result, but instead having failed to produce any result, due to the low acceptability of stimuli. Hence, different types of offline methods were considered instead, as a means for assessing argument status strength in various structures.

## Judgments

Offline methods collect data from participants' final responses rather than during processing. For instance, Koenig et al. (2003) used a sentence completion task to test whether certain types of constituents are more strongly connected to the verbs than other types. Conclusions were drawn from the proportions of various types of information that participants used for completing half-finished sentences.

Another fruitful way of investigating argument status is to use a judgment task. Judgments are a frequently used offline paradigm in experimental approaches in linguistics (Goodall, 2021). Acceptability judgments constitute perhaps their most prominent type, but various other types have been used as well. For instance, Maurer and Koenig (2000) collected contradiction judgments in order to validate their intuitions about a set of verbs requiring agent arguments. In this task, participants were presented with a sentence, followed by another sentence. They had to judge, whether the situation described by the second sentence was logically possible in the context of the first sentence. In addition, sensibility judgments occur in (Boland and Boehm-Jernigan, 1998; Maurer and Koenig, 2000). In this task, participants are tested in a self-paced reading task with an additional judgment element. When they reach a point in a sentence where they feel that the sentence has stopped making sense, they have to indicate it by pressing a button.

Koenig et al. (2003) used a task where participants had to assess whether a verb requires, allows or excludes the existence of certain types of information. For instance, upon assessing the relationship between a verb and an Instrument, a rater had to indicate whether a verb describes a situation in which the Agent must, can, or cannot use an instrument to perform an action. For instance, after reading the sentence 'Marc poked the frog' they had to decide whether using an instrument was optional or not for Marc (Koenig et al., 2003, 79).

This method was developed into an argument judgment task in (Rissman et al., 2015). Argument judgments reflect the participants' opinion on which constituents are the participants in the event or state described by the verb. In this task, experiment participants were first instructed in the canonical ways in which verbs select arguments. For instance, they were told that 'run' is a verb only selecting one argument, because nothing else is required for the event of running than

just a single runner. ‘Eat’, however, describes an event which could not happen without two participants, i.e. the eater and the eatee. After reading instructions, participants were presented with a sentence where two elements were highlighted by brackets, and they were asked to choose which element (or neither) is the argument of the verb. The study also included a second task where participants had to decide what are the arguments of a (bare) verb lemma.

The weakness of this approach is that such judgments rely on argument status, i.e. the concept of various constituents being entailed in the meaning of the event or state. This is not generally a concept with which a naive rater is familiar (contrary to, for instance, the idea of a sentence being acceptable or likely to be uttered). Experiment participants therefore require instruction to ensure some degree of uniformity in the way they understand the rather abstract notion of event participanthood. Instruction, however, risks biasing participants against stimuli.

The requirement of instruction, however, tends to be true of judgment tasks in general (and, one might say, of many other paradigms as well). For instance, with sensibility judgments, participants needed to be instructed in terms of how sentences express insensible content. Hence, prior to the task they were given examples of it (e.g. one cannot wear a house) (Boland and Boehm-Jernigan, 1998, 692).

However, a judgment task is useful as it is less strict about stimuli forms and because it indicates various grades in argumenthood. In a task with binary choice (a participant either judges or does not judge a phrase to be an argument), gradience is reflected by inter-rater agreement. For instance, canonical objects systematically received ‘is argument’ judgments while canonical adjuncts systematically received ‘is not argument’ judgments. This indicates positions in the extremes of the argument status scale. However, for the groups that Rissman et al. (2015) hypothesised to constitute semi-arguments of different strengths, highly diverse participant judgments reflected this property.

### 6.1.3 Semi-arguments

Sentence completion and judgment tasks have been used to demonstrate the non-binary nature of argument status, i.e. that constituents vary in a gradient manner in terms of the strength of the cognitive link between them and verbs (Koenig et al., 2003; Rissman et al., 2015). This variance has been shown to be affected by both the constituent’s semantic type as well as the lexical semantics of the verb.

On the one hand there are indications that elements of a certain semantic category always have stronger argument status than elements of a different category. For instance, Koenig et al. (2003) showed that participant locations, i.e. locations only describing one participant of the event (‘I carried my watch **in my pocket**’) are stronger arguments than event locations describing all participants of the event (‘I painted the picture **in my studio**’). Similarly, Rissman et al. (2015) demonstrated that the extremes of the gradient are occupied by canonical objects and

adjuncts — Themes were frequently selected as arguments while Time, Manner and event locations were selected infrequently.

On the other hand, the argument status of phrases bearing the same semantic role appears to vary. This seems to depend on the lexical semantics of verbs. For instance, a distinction was shown between verbs describing events which require Instruments ('cut') and those which allow them ('break'). The former were judged to select Instruments as arguments more frequently than the latter (Rissman et al., 2015). However, semantic category effects nevertheless persisted in this study, because even the strongest Instrument arguments were selected as arguments by fewer speakers than Recipients for verbs which require them ('send').

This also illustrates that encoding an argument and semantically requiring the presence of a participant are somewhat separate phenomena. Arguably, a cutting event needs an Instrument as strongly as a sending event needs a Recipient, but Recipients were nevertheless judged as arguments more frequently than Instruments (see Rissman et al., 2015, who also demonstrate that these results are not due to animacy effects). In the same sense, many cutting events presumably require Time or a Location, but the constituents expressing them were nevertheless rated as arguments very infrequently. Hence, in addition to the require-allow distinction, argument status also reflects something we might call prominence. The way in which sending events encode Recipients is different from the way in which cutting events encode Instruments. Verbs encode information with various degrees of prominence.<sup>34</sup>

## 6.2 Argument judgment task — aims and research questions

An argument judgment task was organised to investigate gradience in argument status, and the link between argument status and various argument structures. At this point, no distinction is made between semantic and syntactic arguments in terms of argument connection quality. Both are viewed as marking a connection of essentially the same type. The study is built to address three separate gaps in linguistic knowledge.

First, the argument status of oblique arguments has not been systematically investigated in experimental approaches, even though much is assumed about it in linguistic theories. In Estonian linguistics, the argument status of oblique and canonical arguments is assumed to be fundamentally distinct (Erelt, 2003; Veismann et al., 2017) (see chapter 3). Up to now, this assumption has been based on the morphosyntactic properties of these structures, which were discussed in section 2.2.2 and 3.6.1 as highly problematic when used for this purpose.

---

<sup>34</sup> Koenig et al. (2003) argue that differently from adjunct roles, arguments are defined by their particularity — i.e. the fact that they only describe the aspects of *some* verbs. For instance, event location occurs with a much wider range of verbs than Instruments, which only occur with a small set of verbs.

Second, mapping the argument status of oblique arguments allows us to investigate the link between argument status and transitivity. Namely, as was shown in section 4.2.1, verbs selecting oblique arguments can broadly be described as less transitive than verbs selecting canonical objects. This tendency has, however, been used to argue that oblique structures mark a qualitatively more poor type of argument status than canonically marked arguments: “[...] verbs taking peripheral complements are not activity verbs involving impingement on a patient, though this is not to imply that all verbs with accusative-marked direct objects are such verbs” (Blake, 2001, 34). Hence, one needs to explicitly demonstrate that oblique arguments are as strongly connected to verbs as canonical objects, meaning that argument status is distinct from the lexical-semantic transitivity of verbs.

Third, the design of the study allows us to investigate which properties of arguments indicate weaker argument status, offering concrete evidence in support of the hypothesis that argument status is an inherently gradient phenomenon (following Koenig et al., 2003; Rissman et al., 2015). Gradient grammar concepts have been criticised for offering a vague rather than an organised view of gradience and not being grounded in a theoretical framework (Harris, 1993). Hence, instead of merely demonstrating that gradience exists, this study attempts to offer explanations as to which factors affect argument status strength by showing which types of elements occupy different regions of the scale. Hence, the design incorporates various types of phrases marked in spatial cases, which are hypothesised to describe different regions of the scale. In addition to non-spatial oblique arguments, the study also includes metaphorically spatial oblique arguments (section 6.2.1), as well as two types of participant locations (section 6.2.2). All four sets of experimental items entail spatial cases, allowing the study to investigate whether a difference in case functions renders a difference in argument status.

### 6.2.1 Oblique arguments

Verbs in this group originate from the study on the lexical semantics of verbs (chapter 4). However, that sample was somewhat semantically varied. Some encode entirely non-spatial arguments (Themes) while others express some degree of spatial meaning in a metaphorical sense (abstract Goals, Sources, Locations). These two types form two distinct experimental item categories in the present study.

Verbs marking oblique arguments in entirely bleached cases include *sõltuma* ‘depend on’ and *hoolima* ‘care about’. In these instances, the original semantics of the case (in this instance, spatial meaning) is no longer present, not even metaphorically. The case merely marks event or state participation.

Verbs occurring with abstract spatial oblique arguments include verbs which were seen to alternate between concrete and abstract spatial constituents, but occur with abstract arguments more than with concrete locations in corpus data. They include *lähenema* ‘approach’ selecting allative (e.g. *lähene-n poe-le* approach-

PRS. 1SG shop-ALL ‘I am approaching the shop’ vs *lähene-n probleemi-le* approach-PRS. 1SG problem-ALL ‘I am approaching the problem’). The group also includes *investeerima* ‘invest in’ which always encodes spatial meaning. All in all, verbs in this group may be regarded as selecting oblique arguments, even though they also encode spatial semantics in a metaphorical sense.

### 6.2.2 Participant locations

Participant locations have been shown to constitute stronger arguments than event locations in English (Koenig et al., 2003). The two types differ in terms of scope, participant locations only pertaining to one of the event participants (‘I hid the watch **in my pocket.**’) while event locations refer to all participants (‘I ate chocolate **in the garden.**’). The stronger argument status of participant locations is thought to derive from the fact that they occur with a much more narrow group of verbs than event locations (Koenig et al., 2003).

However, the verb lists used in (Koenig et al., 2003) show that another confounding variable may have affected the results. Namely, in addition to being distinct in terms of scope, the two types of locations used in that study also differed in terms of a directional component. A vast majority of participant location verbs in (Koenig et al., 2003) encoded Goals (‘hang’) while a minority encoded stative Locations (‘boil’). Event locations, however, all constituted stative Locations. Hence, it remains unclear whether experiment participants used participant locations more frequently than event locations for finishing sentences because of their narrower scope (and being less prominent across verbs) or because of their directionality, which may well contribute to a stronger cognitive link between the verb and the location.

The present study investigates this by testing two types of participant locations: directional Goals (‘I hid it **in my pocket.**’) and non-directional Locations (‘I washed it **in a sink.**’). Both only describe the location of the object and not the subject, but the former entails a directional component while the latter does not. The study is therefore able to test whether directionality plays a role in argument status strength. See Taremaa (2017) for a comprehensive study on motion (including directionality) in Estonian verbs.

### 6.2.3 Research questions

Altogether, the present study answers two research questions:

RQ7: Does abstract spatial meaning in verbal semantics indicate weaker argument status?

RQ8: Does directionality indicate stronger argument status for physical locations?

We hypothesise that the answer to RQ7 is affirmative, meaning that bleached oblique arguments are expected to constitute stronger arguments than oblique arguments with metaphorically spatial meaning. We also hypothesise an affirmative answer to RQ8, meaning that directional participant locations (Goals) are expected to be stronger arguments than non-directional participant locations.

## 6.3 Experiment method

The argument judgment task presented in this chapter uses a design rather analogous to that of the experiment in (Rissman et al., 2015), in which regards the pre-experiment instructions, the presentation of stimuli and the basic condition system behind stimuli. All data are freely available (Aigro, 2022a).

### 6.3.1 Participants

Participants included 42 adults (11 male and 31 female, mean age 26.9 years, range = 19–54) who completed the task at the Phonetics Laboratory of the University of Tartu. All were native Estonian speakers. Four participants reported having grown up while regularly speaking another language in a questionnaire filled before the task. One participant reported having taken linguistics courses at the university level. Neither group showed any distinct behaviour in their results. Participants received 5 euros for their time. The experiment took 30–40 minutes.

### 6.3.2 Material

As explained in section 6.2, the experiment investigates the argument status of four different types of phrases marked in spatial cases. These four types of phrases co-occur with four different groups of verbs. These verb sets are described in Table 16 together with the fillers.

**Table 16:** Experiment stimuli and control trials (see verb lists in Appendix 2).

	Type	No of verbs	Total trial sentences
1	Bleached oblique arguments	8	48
2	Spatial oblique arguments	8	48
3	Directional participant locations	8	40
4	Non-directional participant locations	8	40
5	Filler trials	15	90
	<b>Total</b>	<b>47</b>	<b>266</b>

All participants saw all trials ( $n = 266$ ). Each oblique verb and control verb occurred in six different trials and each participant location verb in five different trials.

### Bleached and spatial oblique arguments

Bleached and spatial oblique argument verbs (groups 1 and 2 in Table 16) originate from the oblique verb dataset described in section 4.1.3. The experiment includes 8 bleached oblique argument verbs, including those selecting elative ( $n = 3$ , *sõltuma* ‘depend on’, see (64)), allative ( $n = 2$ , *reageerima* ‘react to’) and illative arguments ( $n = 3$ , *investeerima* ‘invest in’).

Differently from bleached oblique argument verbs, spatial oblique argument verbs always encode spatial semantics. The set includes 8 verbs, some of which select elative arguments ( $n = 3$ , *hoiduma* ‘keep away from’), some allative ( $n = 2$ , *lähenema* ‘approach’) and some illative arguments ( $n = 2$ , *sukelduma* ‘dive into’). As mentioned above, the group only includes verbs, which were confirmed in section 4.1.3 to select abstract complements more often than concrete complements in corpus data. In the trials of the present experiment, they were only presented with abstract complements (*lähene-b probleemi-le* approach-PRS.1SG problem-ALL ‘approached a problem’).

Hence, both sets include eight verbs, each of which occurs in six sentences. Each set is therefore represented by 48 sentences.

### Directional and non-directional participant locations (Group 3–4)

Participant location verbs in groups 3 and 4 (Table 16) mostly constitute Estonian translations of the participant location verbs used in (Koenig et al., 2003). All English participant location verbs in (Koenig et al., 2003) ( $n = 36$ ) were coded for semantic role. Seven verbs with Goals were translated into Estonian and added to group 3 (*peitma* ‘hide’, *valama* ‘pour’). Another verb was added to the set (*asetama* ‘set’/‘put’/‘position’).

Six participant location verbs in (Koenig et al., 2003) coded as selecting Locations were added to group 4 (*grillima* ‘grill’, *reklaamima* ‘advertise’). Two other

verbs were added to the set: *pesema* ‘wash’ (e.g. *pesi-n särki kraanikausi-s* approach-PST.1SG shirt.PAR sink- approach-PRS.1SG shop-INE ‘I washed the shirt **in a sink**’) and (*kasvatama* ‘grow’, e.g. *kasvata-n peenra-l tulpe* grow-PRS.1SG flower.bed-ADE tulip.PAR.PL ‘I am growing tulips **on the flower bed.**’).

This resulted in two datasets, both of which contained 8 verbs, each totalling 48 trial sentences. However, differently from oblique arguments, participant locations occur as the third argument rather than the second. Hence, these trials also included a direct object, including five distinct segments rather than four. This is discussed below.

### Filler trials

15 verbs were included in the filler set, all of which select canonical direct objects. They serve several roles in this study. First, they were intended to function as fillers, diversifying the set of sentences seen by each participant. Second, they allow the study to check on how well participants understood the task. This means that direct objects are expected to be somewhat frequent argument choices. Third, they were to serve as a baseline category, providing information on how often canonical objects are regarded as arguments and how they compare to oblique arguments.

This is why filler sentences were presented in the same condition system as experimental items (see section 6.3.3), with a structure similar to that of oblique argument trials (discussed in section 6.3.3).

In addition, the set includes verbs with various degrees of transitivity, so they would be at least somewhat representative of canonical object verbs in general. Hence, the group includes verbs depicting dynamic events ( $n = 12$ ) as well as non-dynamic events ( $n = 3$ ). It also includes those expressing telic events ( $n = 7$ ) and atelic events ( $n = 8$ ). Some describe events where objects are physically affected (*parandama* ‘fix’, *lõhkuma* ‘break’) while in others objects remain unaffected (*mõõdta* ‘measure’, *tõlkima* ‘translate’).

Each of the 15 filler verbs occurred in 6 sentences, five of which included a direct object. Of the 75 different object nominals, 42 included an object marked in partitive case, 13 had a genitive object and one had a nominative object. In 19 sentences, object case was ambiguous between partitive and genitive.

### 6.3.3 Trials

During the task, participants saw sentences in Estonian where two phrases were presented in brackets, as in (64):

- (64) Andres hoolis [loodusest] [tuli hingeliselt]  
 Andres cared [about nature] [fiercely]

Their task was to decide which of the two bracketed elements constitutes an argument of the verb. They could also choose the option that neither is an argument.

All bleached and spatial oblique argument trials as well as filler trials included 4 lexemes (64). Verbs occurring with participant locations, however, were presented in trials containing 5 lexemes (65). This is because these verbs encode canonical objects in addition to participant locations.

- (65) Mart riputas *midagi* [rõdule] [uniselt]  
Mart hung *something* [onto the balcony] sleepily

In order to control for the potential effect of the NP's position in relation to the verb, all trials had two versions. Half of the participants saw a version where the target phrase immediately following the verb (64) while the other half saw a sentence where the modifier occurred between the target phrase and the verb (66):

- (66) Andres hoolis [tuliingeliselt] [loodusest]  
Andres cared [fiercely] [about nature]

Half of the sentences seen by each participant followed one word order pattern and half followed the other. The order of sentences was randomised for each participant.

### Contrast types

As mentioned above, each oblique and filler verb occurred in six different sentence contexts, while each participant location verb occurred in five. Following Rissman et al. (2015), half of these contexts were designed to contrast the target phrase (oblique argument, participant location) with various types of modifier phrases. The aim of such a setting is to show that the target phrase would be selected as an argument more often than the modifier occurring in the same sentence. The second half of the trials lack target phrases altogether, pitting modifiers against each other. The aim is to show that when presented with a choice between two modifiers, neither would be chosen as an argument.

Hence, all 16 oblique argument verbs occurred in 6 different trials, each representing a distinct context. Each contain a subject and a verb, but they vary in terms of the phrases pitted against each other for argument judgments. These combinations are:

1. Oblique argument vs event location
2. Oblique argument vs manner
3. Oblique argument vs time
4. Event location vs manner
5. Event location vs time
6. Manner vs time

Each participant location verb (n = 16) occurred in five different contexts, because they were kept from co-occurring with event locations. Hence, the target phrase (Part\_loc) was pitted against a modifier in two of the five trials per each verb:

1. Part\_loc vs manner
2. Part\_loc vs time
3. Event location vs manner
4. Event location vs time
5. Manner vs time

In filler trials, the adverbial phrases were even more diverse, in order to bring variation into the overall set of trials. Each filler verb occurred in the following six contexts:

1. Object vs location
2. Object vs beneficiary
3. Object vs amount
4. Object vs purpose
5. Object vs instrument
6. Modifier vs modifier

The 30 modifier phrases included various types of adverbials, chosen based on verb semantics. Most were event locations (37%), but the set also includes comitative co-participants (25%), time (19%), manner (13%), instruments and material (3% each). Half of these phrases were made up of one lexeme and half by two.

Various modifier phrases occurring with oblique argument, participant location and direct object verbs were kept structurally uniform throughout. In experimental trials, all phrases contained just one lexeme, except time phrases, which contained two (*eile õhtul* ‘yesterday evening’) and event locations which alternated between one and two. In control trials, all non-object phrases were made up of two lexemes, except for purpose phrases, which contained four lexemes (e.g. *et külm ei hakkaks* ‘so they would not get cold’).

All directional participant locations were marked in allative or illative case, while non-directional participant locations were marked in adessive or inessive. Time phrases were marked in inessive or adessive, instruments in comitative, never occurring with the adposition *koos* ‘with’. Beneficiary phrases ended with the postposition *jaoks* ‘for’. Manner adverbs were formed with the *-lt* derivation (*õnnelikult* ‘happily’). Subjects were human and singular, containing both common human nouns as well as proper names. No phrase occurred twice in any two trial sentences.

Finally, as explained in section 6.1.1, there are theories which take a frequentist view on argument status, viewing it as something that arises from the frequent co-occurrence of verbs and arguments, rather than a special syntactic/lexical status of the latter. In order to make sure that the effects observed in the present study are not due to co-occurrence frequency distinctions, all experimental trials were controlled for collocation frequency (log-likelihood) between the verb and the target phrase (oblique arguments and participant locations). Hence, no verb occurs with less common target phrase than other verbs.<sup>35</sup> In addition, verb lemma frequencies were also kept similar between groups.

### 6.3.4 Procedure

Before the task began, each participant read through on-screen instructions describing the basic ideas behind the concept of argument status. Instructions were designed in an analogous way to those used in Rissman et al. (2015). Instructions are available in the data repository of this work (Aigro, 2022a).

Instructions presented participants with four principles about argumenthood. First, participants were invited to consider arguments as elements which are necessary for the verbs to have meaning and the events or states described by the verbs to happen. Two example sentences were provided, which both were presented as having two arguments: *kallistama* ‘hug’, which requires that there be a hugger and someone that is hugged, and *tahtma* ‘want’, which requires that there is someone that wants and something that is wanted.

Second, participants were told that arguments could occur in the same sentence as the verb, as in *Poiss taha-b jäätis-t* boy.NOM want-PRS.3SG ice.cream-PAR ‘The boy wants ice cream’. However, they can also be omitted from the sentence, as in *Anneli-t kallista-ti* Anneli-PAR hug-IMPRS.PST ‘Anneli was hugged.’ In any case, the nature of arguments encoded by these verbs is unaffected by this.

Third, participants were told that verbs differ in terms of how many arguments they take. While ‘hug’ selects two arguments, *jooksma* ‘run’ (*Jaagup jooksi-s* Jaagup.NOM run-PST.3SG ‘Jaagup ran.’) only selects one. This is because nothing but the runner is required for a running event to take place.

Finally, participants were told that while sentences can include all types of phrases, not all of them are arguments. They were given two examples where only one argument is present: *Jaagup jooksis, kuni tal hakkas paha* ‘Jaagup ran until he was sick’ and *Jaagup jooksis, sest zombi ajas teda taga* ‘Jaagup ran, because a zombie was chasing him’. The instruction highlighted the fact that because Jaagup can run without reaching any particular physical sensation, and also without knowing why he runs, these two phrases are not arguments of *jooksma* ‘run’. However, phrases expressing the reason or duration of events can still be the arguments of other verbs.

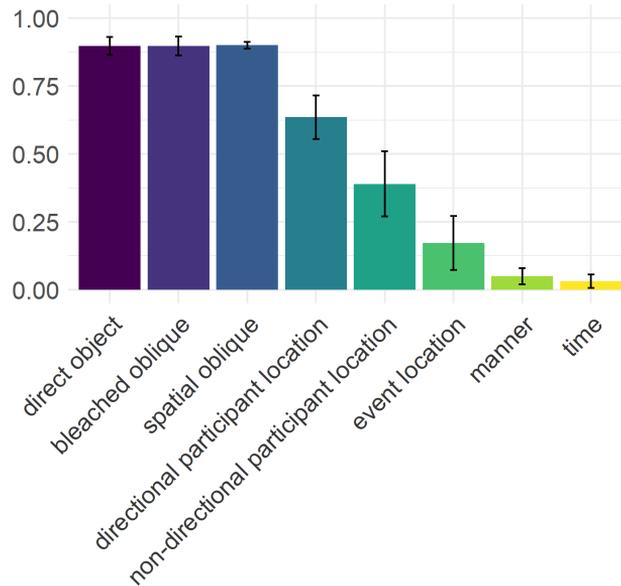
---

<sup>35</sup> Log-likelihood data originates from the database of collocation information from the Estonian National Corpus: <https://korpused.keeleressursid.ee/clc/?mod=search>.

After the instructions, participants were encouraged to ask the experimenter questions about what they learned. After this, the task started. Participants saw a sentence on the screen (e.g. (64)), with the first bracketed phrase being numbered as [1] and the second as [2]. They were instructed to press [1], [2] or [3] on the keyboard, indicating their choice: [1] is an argument, [2] is an argument, or [3] for marking that neither is an argument. The task was compiled with Experiment Builder.

## 6.4 Results

Fig. 26 presents an overview of the proportion in which the four types of experimental target phrases were chosen as arguments of the verb with which they co-occurred, mapping them with objects from filler trials and three types of modifiers:



**Figure 26:** Proportion of selecting various constituents as arguments of a verb, including target phrases from experimental trials (bleached oblique argument, spatial oblique argument, directional participant location and non-directional participant location) and filler trials (direct object), as well as various modifier contrasts (event location, manner, time).

First, the results confirm that the extremes of the argument status scale broadly align with canonical objects and canonical modifiers. In filler trials where canonical objects are contrasted with various modifiers (amount, beneficiary, location, instrument and purpose phrases), objects were selected as arguments in 89.7% of instances. Between the 15 verbs in the filler set, the proportion of choosing ob-

jects as arguments did not vary greatly, remaining between 88–92% for all verbs except *laulma* ‘sing’ (80.5%), *koristama* ‘tidy’ (85.2%) and *parandama* ‘mend’ (94.2%).

Modifiers were generally only rarely chosen as arguments. First, in experimental trials where one modifier was contrasted with another, the option ‘neither is an argument’ was chosen in the majority of trials (80.2%). This shows that the design worked in the sense that participants did not feel obliged to choose one of the constituents as arguments no matter their type. However, there was some variation among canonical modifiers. While time and manner phrases were chosen as arguments very infrequently (2.6% and 4.0%), event locations were seen as somewhat more argument-like, being chosen in 13.1% of sentences (Fig. 26).

#### 6.4.1 Argument status of elements marked in spatial cases

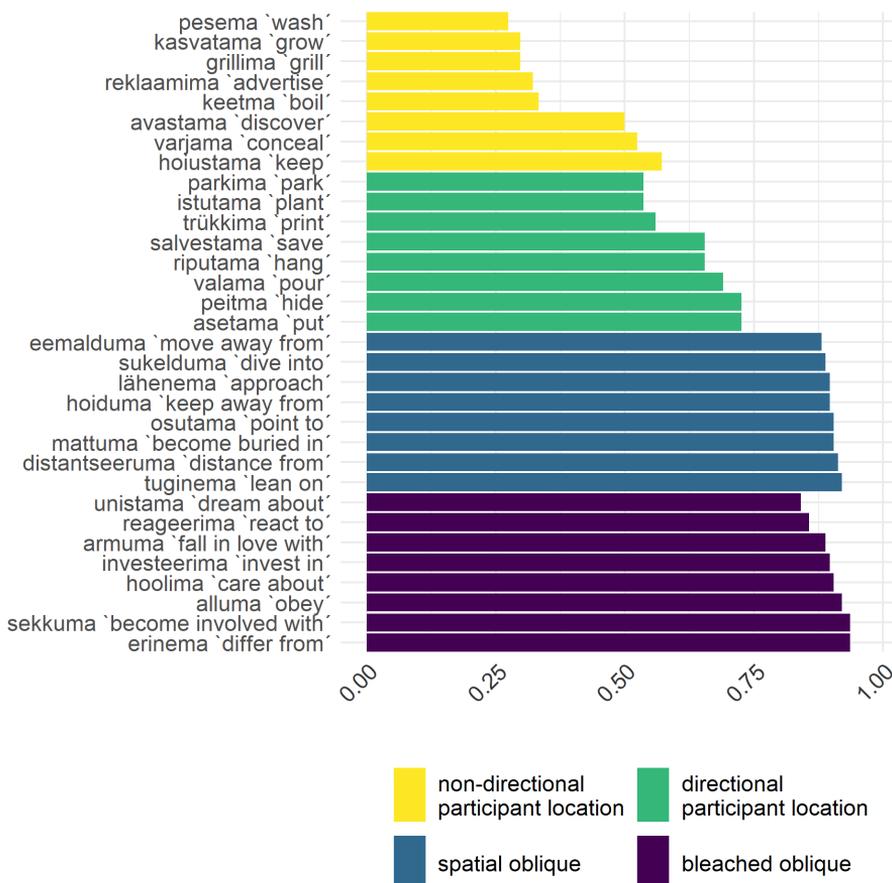
As seen in Fig. 26, entirely bleached oblique arguments (e.g. the relative argument in *sõltu-b kasumi-st* depend-PRS.3SG profit-ELA ‘depends **on the profit**’) are as strong in terms of argument status as canonical objects (*põleta-b paberi-t* burn-PRS.3SG paper-PAR ‘burns paper’): mean: 89.8% and 89.7%, sd: 3.5% and 3.3%, respectively).

However, rather surprisingly, the same applies to spatial oblique arguments (e.g. *lähene-b probleemi-le* approach-PRS.3SG problem-ALL ‘S/he approaches **a problem**’), which are also very similar to direct objects in terms of argument status strength (mean = 90.1%, sd = 1.3%). Metaphorical spatial meaning in verbal lexical semantics does not therefore indicate reduced argumenthood of its arguments, because these verbs pattern similarly to those with no spatial meaning whatsoever, i.e. bleached oblique arguments and direct objects (Fig. 26).

Finally, the argument status of directional participant locations is stronger than that of non-directional participant locations (directional: mean = 63.5%, sd = 8.1%; non-directional: mean = 39.0%, sd = 12%). This indicates that spatial directionality does play a role in terms of the quality of the link between verbs and types of spatial constituents.

## Verb-based analysis

The distinctions found between the four experimental verb groups in Fig. 26 remain in place when investigating individual verbs, as shown in Fig. 27:



**Figure 27:** Argument status strength per verb in the four experimental groups.

There is a rather distinctive break between oblique arguments and participant locations. Verbs in the category ‘spatial oblique’ in Fig. 27 were all presented with abstract complements, e.g. *lähene-b probleemi-le* approach-PRS.3SG problem-ALL ‘S/he approaches the problem’. Abstract and concrete spatial phrases are clearly judged as distinct (‘spatial oblique’ and both types of participant locations in Fig. 27). The same distinction is not made between abstract spatial semantics and the lack of spatial semantics (‘spatial oblique’ and ‘bleached oblique’ in Fig. 27). The border between the two participant location groups is slightly more gradient, both referring to concrete locations.

## Mixed-effects logistic regression

We fit a mixed-effects logistic regression model to test the statistical significance of the distinctions found between the four experimental stimuli groups. We used the lme4 package in R for the model (Bates and Maechler, 2009). Data for the model only includes trials where target phrases were pitted against modifiers, excluding trials where modifiers were pitted against modifiers. The model includes one dependent variable (judgment type, describing whether or not the target phrase was chosen as the argument) and two independent variables (trial type, collocation frequency between the verb and target phrase), as explained in Table 17:

**Table 17:** Variables included in a mixed-effects logistic regression model. The model includes three random effects: *participant*, *verb lemma* and *sentence*.

Variable	Type	Levels
Judgment type	Categorical	‘target phrase’ and ‘non-target phrase’
Trial type	Categorical	‘bleached oblique argument’ ‘spatial oblique argument’ ‘directional participant location’ and ‘non-directional participant location’
Collocation frequency	Numerical	The log-likelihood function of collocation <sup>36</sup>

Table 18 presents the results of the model:

**Table 18:** Results of logistic mixed-effects regression

Condition	Estimate ( $\beta$ )	Standard error	P value
Intercept	0.768220	0.323958	0.0177*
Trial type = bleached oblique	2.272724	0.256068	< 2e-16***
Trial type = spatial oblique	2.294781	0.257142	< 2e-16***
Trial type = non-directional participant location	-1.449086	0.244788	3.22e-09***
Collocation frequency	-0.000305	0.004622	0.9474

The Intercept describes a hypothetical situation where directional participant locations are selected as arguments. Selecting the target phrase as an argument is significantly more likely with both spatial and non-spatial oblique argument verbs

than with directional participant locations ( $p < 0.0001$ ). Selecting a bleached oblique complement as an argument is 9.7 times more likely, and a spatial oblique complement 9.9 times more likely than selecting a directional participant location as an argument.

With directional participant locations, choosing the target phrase is 2.2 times more likely than choosing a modifier as an argument. With non-directional participant locations, however, choosing other modifiers is slightly more likely than choosing participant locations.

Importantly, collocation frequencies have no significant effect on the choice of argument in this study. This means that the effects observed in the results of this study cannot be explained by the verbs in some groups co-occurring with more typical or natural complement or participant location phrases.

## 6.5 Summary

This study aimed to answer two RQs: Does abstract spatial meaning in verbal semantics indicate weaker argument status (RQ7)? Does directionality indicate stronger argument status for physical locations (RQ8)?

RQ7 was received a somewhat unexpected answer. Bleached oblique complements are strong arguments as was expected, scoring very similarly to canonical objects. Unexpectedly, however, so are (abstract) spatial oblique complements. It appears that abstract spatial meaning does not lead to weaker argument status as participants broadly rated the argument status of an abstract complement of *lähenema* ‘approach’ as strong an argument as the complements of verbs such as *sõltuma* ‘depend on’, as well as canonical objects.

For RQ8, however, the main hypothesis was confirmed to apply. Directionality plays a role in argumenthood, meaning that directional semantics in verbal lexical semantics increases the strength of its co-occurring participant locations’ argument status. Concrete spatial Goals have a stronger cognitive link to the verb than concrete spatial Locations.

In addition, results also suggest that location scope also plays a role, as was assumed in (Koenig et al., 2003). Locations with a more narrow scope (participant locations) are stronger arguments than locations with a wider scope (event locations). However, this indication would require further proof as these two groups were not compared in an equal manner in the present study.

The somewhat higher argument status of event locations, compared to that of time and manner adverbials, is likely an artefact of the design of the study. Given the approach, the stimuli verbs might make a stronger reference to all constituents with spatial semantics. This is also evident from verb group analysis, where verbs selecting non-directional participant locations are especially likely to have event locations chosen as arguments (15.9%, compared to 10.5% in bleached oblique argument trials). Indeed, in such sentences the participant location ((67a)) is very similar to event location ((67b)):

- (67) a. Väikeettevõtja reklaamis midagi [koduleheküljel] [kõlavalt]  
 An entrepreneur advertised *something* [on a website] [convincingly]  
 b. Sõbranna reklaamis midagi [peol] [reede öösel]  
 Girlfriend advertised *something* [at a party] [Friday night]

Each research question was answered by comparing two groups of sentences, which were presented to the participants in an analogous manner. The two sets of trials (oblique argument trials and participant location trials) differed from each other in that participant location trials were longer in terms of lexeme count, presenting the target phrase as a potential third rather than a second argument. This is why one cannot clearly draw a conclusion about how oblique arguments compare to participant locations in terms of argument status. However, the study does propose a strong hypothesis in this regard. All oblique arguments in the study were judged to constitute arguments more often than all physical locations. Hence, these results suggest that oblique arguments have stronger argument status than physical locations. This claim would require proof in further studies, where oblique arguments are compared to locations in a similar syntactic position, e.g. using one-place verbs where locations occur in the second position.

These findings also have a number of implications for the nature of lexical argument structure, its relationship with transitivity and the way different types of elements occupy the argument gradient.

### 6.5.1 Gradience in argumenthood

First, these results support the hypothesis that argumenthood is a gradient property of various constituents describing event-related information. We assume there are various semantic factors that facilitate the strength of argument status. This study examined the spatial axis by focusing on various elements where spatial cases occur in different functions and mark different degrees of argument status.

In this view, oblique arguments marked in spatial cases that lack any cohesive semantics (cases occurring in bleached function) are in one extreme of the scale. Together with canonical objects and spatial oblique arguments they constitute strong arguments. This means that verbs are strongly linked to spatial case marked constituents expressing event/state participants, where case does not mark concrete physical space. When a case does mark physical space, argument status is weaker. Loss of directionality was shown to facilitate this decrease, but results also suggest that a wider location scope indicates weaker argument status (Koenig et al., 2003). Directional participant locations actually constitute semi-arguments.

The results of this study also imply that transitivity is not a reliable indicator of argument status. Section 4.2.1 showed that transitivity affects argument structure assignment. Most spatial case argument structures are more likely to occur with low-transitivity verbs than the canonical structure. Results of the present experiment, however, show that even states with very low transitivity, e.g. *sõltuma* ‘de-

pend on' and *hoolima* 'care about' encode strong arguments. All bleached oblique argument verbs in the stimuli can be said to have been broadly less transitive than the verbs selecting canonical objects, but the two groups showed no distinction in argument status strength. In fact, the highest argument score among oblique arguments belongs to the argument of the highly intransitive abstract property verb selecting elative case, *erinema* 'differ from' (93.7%). For instance, the highest argument score in the direct object group belongs to the highly transitive *parandama* 'mend' (94.3%). Hence, it appears there is no direct link between transitivity and argument status. However, transitivity could nevertheless affect argument status in a more complex way in particular types of verbs or clauses.

### 6.5.2 Factors affecting argument status

This chapter aimed outline a few factors that affect argument status strength. While future studies are needed (in other languages) to confirm them, it nevertheless found indications of several such factors along the spatial dimension:

1. The directional component of spatial elements affects argument status. This was evident from the fact that directional participant locations were selected as arguments significantly more than non-directional participant locations.
2. The scope of location may affect argumenthood, as was claimed by (Koenig et al., 2003). This means that participant locations which only describe the location (or destination) of one of the event participants, constitute stronger arguments than event locations, which describe all event participants simultaneously. This claim needs further proof in future studies as the present study merely suggests this as a plausible hypothesis.
3. Concreteness of arguments may affect argumenthood. Contrary to our assumption that metaphorical spatial semantics indicates weaker argument status on its arguments, spatial oblique arguments (with abstract referents) were shown to constitute strong arguments. The argument status distinction between participant locations and oblique arguments may be rooted in their concreteness distinction. However, it could also stem from the fact that participant locations occurred as third arguments (following canonical objects), while oblique arguments occurred in the second place. Hence, further studies are required, where the number of participants and concreteness are manipulated among stimuli.

## 7 DISCUSSION

Three broad aims are set for this thesis in chapter 1 — 1) to investigate the distinctions between verbs in oblique and canonical argument structures in terms of lexical semantics and usage, 2) to link the properties and development of individual cases to the system of argument structures and 3) to investigate the argument status of oblique arguments and outline the argumenthood gradient by means of comparing the argument status of various constituents in spatial cases.

This thesis presents the results of five corpus studies, one experiment and one study simultaneously using corpus and behavioural data, to address these three topics. All datasets are available to be used freely (Aigro, 2022b,a). The following will discuss the main results. First, I outline what was learned about the bleaching of spatial cases in Estonian and their emergence as argument structures (Aim 2). Next, section 7.2 discusses the argument structure system in Estonian, focusing on the ways in which structures are sensitive to verbal lexical semantics, as well as how they are used (Aim 1). It also discusses the results on the argument status marked by oblique structures, discussing the gradience of argumenthood in the context of frameworks used for describing gradient grammatical phenomena (Aim 3).

Finally, section 7.3 discusses the implications of the present results on the discussion on *why* a language might have such a varied argument structure system in the first place.

As discussed in section 2.1, the present findings are relevant to both lexicalist and constructionist approaches. The thesis constitutes a range of empirical findings, which are presented in a rather a-theoretical way. This is because more data is needed for both lexicalist and constructionist approaches to be able to review the ways in which different elements have been conceptualised. Any theory ought to be able to explain present findings. Canonical typology is one great example of this, because as is discussed in section 7.2.4, the ways in which studies using this framework have tackled argument status are rather problematic. Section 7.2.4 discusses more productive future directions for working with arguments and argument marking in Canonical typology. Research such as the present thesis constitute the necessary empirical work required for assessing different approaches and their future directions.

### 7.1 The diachronic perspective

Several aspects of case evolution and grammaticalisation are studied in this thesis by means of synchronic corpus studies. Based on the data, this thesis provides a sketch on how spatial meaning became bleached in Estonian spatial cases and how their usage became extended. Six phases were outlined for the process of bleaching (see Table 14):

Phase 1: concrete spatial meaning (*tooli-l* chair-ADE ‘on the chair’)

Phase 2: abstract spatial meaning (*investeeri-n aktsia-te-sse* invest-PRS.1SG stock-PL-ELA ‘I am investing in stock’)

Phase 3: metaphorically spatial, coherent meaning (*esmaspäeva-l* Monday-ADE ‘on Monday’, *isa-le* dad-ALL ‘to dad’)

Phase 4: non-spatial, coherent meaning (*räägi-n töö-st* talk-PRS.1SG work-ELA ‘I am talking about work’)

Phase 5: non-spatial, non-coherent event/state participant, unaffected by the event (*sõltu-b ilma-st* depend-PRS.3SG weather-ELA ‘depends on the weather’)

Phase 6: non-spatial, non-coherent event/state participant, affected by the event (*paranda-n raadio* fix-PRS.1SG radio.GEN ‘I am fixing the radio’)

The phases describe how the rich case system of a language may give rise to a rich argument structure system. Cases start occurring in new contexts and with a wider range of nominals, their meanings become bleached and less coherent, giving rise to new phases where case usage is even less coherent. Cases become reanalysed as parts of argument structures around phases 3–4. At this stage, their grammaticalisation has reached a point where they are not only fixed on nominals, but they are also syntactically fixed to other elements (verbs). They now constitute patterns that may be assigned to new verbs.

All Estonian spatial cases were shown to be functional in phases 1, 2, 3 and 4, but five out of six spatial cases also feature in phase 5, i.e. they are used as parts of argument structures. Phase 6 was not detected among Estonian oblique argument structures, but it was shown that structural cases (partitive, genitive, nominative) do rarely mark this function with verbs selecting canonical objects (chapter 4).

Even though Estonian spatial cases technically have the functionality of five phases, they differ in terms of how frequently they occur in them. For instance, inessive and illative mostly occur in the functions described by phases 1 and 2. Elative, however, marks these functions only rarely, occurring in phase 5 functions much more frequently.

This illustrates that instead of swapping some functions for others, Estonian spatial cases appear to *broaden* their functionality instead. This means they retain both functions representing earlier phases as well as newer, more grammaticalised functions. Instead of having become structural cases, they have become markers with broad and mixed functionality. This is interesting as it raises questions about the nature of case grammaticalisation and the validity of the idea that grammaticalisation represents a teleological process (Nichols and Timberlake, 1991).

Elative and allative were highlighted as especially functionally mixed, because approximately half of their instances no longer bear spatial semantics. Their argument structures are also more productive than those of other spatial cases,

meaning they are the most likely to become extended to novel verbs. However, inessive and adessive structures may well be extended to novel verbs as well, if the semantics of the novel verb matches that of the few verbs they currently mark as an analogy-based extension.

In addition, these results show that when it comes to investigating the way cases are used in grammar, focusing on merely a few of their functions only tells part of the story. Cases may have a wide range of different functions, which remain systematically connected to each other. Explaining the use of case in grammar must entail descriptions of their overall functionality, because it affects all domains in which these cases appear. For instance, the finding that elative is one of the most grammaticalised spatial cases explains the finding that elative argument structure is used most similarly to the canonical argument structure.

The idea that function frequencies matter when it comes to outlining case usage in structural function, was already expressed by Kuryłowicz (1964, 179): “[...] all case-forms share both kinds of functions, but the syntactical function is *primary* with “grammatical” and *secondary* with “concrete” cases [...]”. Hence, a case may cross over from semantic to syntactic when its function proportions change, but it may also linger on between the two categories, marking both concrete spatial semantics and argument relations, as well as the various degrees between the two extremes. Future typological studies might therefore benefit from an objective system for gradiently quantifying multifunctionality in case systems (as was suggested by Nichols, 1983). This thesis proposes three measures that are useful for such a quantification (functional profiles, lexical variance, syntactic productivity), but as it is not a cross-linguistic study, it does not build a ready-to-use system for such studies.

## 7.2 The synchronic perspective

This thesis studied verb distribution in argument structures and the usage of these verbs (Aim 1), as well as the argument status in different structures and its gradient nature (Aim 3). The following two sections discuss the findings in these topics.

### 7.2.1 Verbs and argument structures

This thesis offers support to widespread hypothesis that oblique argument structures tend to be assigned to verbs with increased stativity and therefore decreased transitivity (Tsunoda, 1981; Malchukov, 2005). The fact that very little dynamicity was found among verbs with non-canonical argument structures indicates that a lack of dynamicity facilitates the assignment of a non-canonical argument structure to novel verbs. The two types of verbs that appear to be especially drawn to oblique argument structures are (1) non-event states (K-states) expressing properties of entities or relationships between two entities (*sõltuma* ‘depend on’) and (2) non-dynamic events (*andestama* ‘forgive’). Lexical dynamicity plays a more

important role in Estonian argument realisation than lexical aspect (contrary to the assumptions made by the aspectual approach to argument realisation, see Tenny, 1992).

In addition, the quantitative approach undertaken in this thesis showed that contrary to what is thought (Nichols, 1983; Blake, 2001), one should not assume that the original semantics of cases always or often follows them into structural function. The study found that a range of verbs govern spatial cases with no spatial semantics whatsoever (e.g. the allative argument of *alluma* ‘obey’). The higher the degree of grammaticalisation of a case, the less frequently original semantics is represented in its argument structure.

Finally, this thesis also investigated the way argument structures are used. The main finding was that, overall, oblique argument structures were more frequently used for talking about humans than canonical argument structures. This applies even when excluding verbs selecting Recipients and Experiencers from the set, meaning that even oblique Theme and Goal arguments referred to humans more frequently than canonical objects (e.g. the arguments of *armuma* ‘fall in love with’, *alluma* ‘obey’, *hoolima* ‘care about’, etc).

This finding is interesting in the context of the phases which guide the evolution of a semantic case. Aristar (1997) suggested that spatial cases are commonly expected to favour nominals with concrete inanimate referents. However, when they start occurring with human nominals (phases 2 and 3 in the pathway above), this means their overall meaning starts to become more bleached. This is because by combining them, spatial cases gain new types of meaning — i.e. one would not use the allative form of *ema* ‘mother’ (*ema-le* mother-ALL) to refer to the placement of something on top of mother’s head (\*‘I put the hat onto mum’). Instead, this new combination of a spatial case affix and a human nominal is commonly interpreted as referring to a more complex and abstract involvement of a participant in some event (e.g. a giving event: *And-si-n raamatu ema-le*. give-PST-1SG book.GEN mother-ALL ‘I gave the book to mum.’). Hence, human nominals are crucial for spatial case grammaticalisation, because the new meanings introduced by them facilitate the further bleaching of these cases.

However, why would oblique structures *exceed* canonical structures in terms of arguments referring to humans? There are several possible explanations to this. One view holds that it is precisely *because* of human nominals and the nature of events involving them that languages have oblique structures in the first place (Lestrade, 2010). Another view is more diachronic and holds that an only somewhat grammaticalised argument structure with a spatial case would be expected to occur with an increased proportion of human nominals. This is because they mark an in-between phase in the development of that case — one that leads to further bleaching (Aristar, 1997). These two perspectives are further discussed in section 7.3.

## 7.2.2 Argument status

The results of this thesis suggest that all argument structures mark the same type of connection between verbs and arguments. The strength of this connection does not depend on the case selected by the verb, nor the degree to which the structure is integrated into syntax (the number of syntactic processes applying to it, e.g. whether or not it may undergo passivisation). Hence, this thesis offers evidence for wider and more inclusive definition for arguments and argument status than what is held by the traditional view, which views canonically marked arguments as distinct from oblique arguments in terms of argument status.

In this more gradient and inclusive perspective, there is no evidence that argument status is linked to particular types of morphemes or syntactic processes, nor to lexical transitivity (the stativity of the verb). Instead, argument status strength appears to depend on various semantic aspects, some of which were shown to play a role in the present thesis. When investigating the weakening of argument status across groups of elements marked in spatial cases, it was found that a lack of concrete spatial meaning in a complement indicates strong argument status. This was somewhat surprising, because it means metaphorical spatial meaning does not lead to weaker arguments (*osuta-si-n probleemi-le* point-PST-1SG problem-ALL ‘I pointed at the problem’). This means the argument class may be somewhat wider than even gradient approaches may assume. Future studies ought to determine if it is the abstractness of arguments that strengthens the argument status of such constituents or their syntactic position as the second rather than the third argument.

In addition, argument strength was shown to depend on whether or not the clause expresses directional semantics, and that it might depend on scope (the number of participants to which a location phrase refers). This, too, ought to be probed in more detail in future studies, where event locations feature as a separate experimental item group, rather than only among contrast modifiers.

All in all, it appears that strong argument status is linked to bleaching and abstractness on the part of complements. One could say that the existence of argument structure (the invariable selection of a case or cases by a verb) implies the existence of strong argument status. However, this phrasing essentially just leads us back to the idea of bleaching and abstractness. This is because one may only hypothesise that something is an argument structure when one frequently observes that a case (or cases) and a range of verbs frequently co-occur in contexts where that case expresses no concrete semantics. In any case, bleaching and abstractness underlie both explanations.

It is important to note that bleaching and abstractness refer to two different phenomena. A case is used in a bleached way when even when the nominal refers to a concrete entity, its combination with the case is not interpreted to have spatial meaning. For instance, the relative complement in *välju-n maja-st* exit-PRS.1SG house-ELA ‘I am exiting the house’ is understood to refer to a spatial relation, but the relative complement in *sõltu-b maja-st* depend-PRS.3SG house-ELA ‘depends

on the house’ is understood to not have spatial meaning. Hence, verbal semantics gives elative a bleached interpretation in the second example but not in the first. Abstractness, however, is a lexical quality. Hence, the event is understood to lack concrete spatial meaning when a spatial case combines with an abstract rather than a concrete nominal (*mattu-n lumme* be.buried-PRS.1SG SNOW.ILL ‘I am buried in snow’ versus *mattu-n probleemi-de-sse* be.buried-PRS.1SG problem-PL-ILL ‘I am buried in problems’). This thesis predicted that bleached interpretation leads to strong argument status and abstract interpretation does not. However, it found that both lead to strong argument status.

In any case, more studies are needed to test these questions in other languages, as well as to probe other factors which potentially might play a role in argument status strength. As these factors appear to be specific to individual semantic domains, more studies are needed to map argument status factors in the spatial domain, but other domains need exploring as well. For instance, the stages of bleaching in the comitative domain was briefly discussed in this thesis, from which potential factors can be derived. After more results have been accumulated in various domains, one might compare their prevalence in the world’s languages and investigate whether some domains are more likely to give rise to argument structures than others.

### 7.2.3 Argument structures as complementary patterns

In the light of these results, it is problematic to view the canonical argument structure as marking fundamentally different argument connection from that marked by oblique argument structures. It does not appear to hold that the degree to which a structure is integrated into syntax (i.e. the number of syntactic processes applying to it, e.g. passivisation), nor the degree to which its cases are grammaticalised (genitive and partitive vs spatial cases in Estonian), nor the semantic transitivity of verbs affect argument status strength. As these factors do stand out in linguistic data, it is understandable that studies based on single examples or small semantically extracted datasets have given reason to assume that argument status does depend on them. The canonical object structure is much more integrated into syntax, dominating data to a rather high degree in some languages (e.g. 88% of Estonian verbs were found to occur with canonical structures). Furthermore, English, the structure of which has unquestionably strongly affected the way linguistic ideas have been conceptualised over the past century, marks oblique structures by means of adpositions. This, too, has contributed to consolidating the idea that oblique structures constitute a peripheral phenomenon entirely distinct from the canonical structure.

However, new quantitative mixed-methods approaches in case-rich languages have the opportunity to cast doubt on this century-old view, introducing quantitative data that pose real problems for that view. They are able to demonstrate similarities between the ways in which structures are used and the way argument

status in them is perceived. They are also able to explain (and sometimes falsify) the discrepancies which traditional approaches emphasise as grounds for making a distinction between structures.

As was said, the idea that canonical argument structures have unique status is somewhat understandable, given its dominance in data and the tendency of example-based approaches to overemphasise the role of frequent strategies. However, this assumed distinction between structures is also unexpected. Such a differentiation is not common in other, yet analogous domains in linguistics. One such domain is derivational morphology. Derivational data has long been explained through the lens of patterns with different degrees of productivity. Just because two deverbal nominalisation suffixes differ in productivity, they are not regarded as yielding as nominalisations of different quality. A morpheme may even strongly dominate a particular domain but assuming qualitative distinctions is uncommon.

For instance, Estonian has two manner adverbial suffixes that may be applied to various types of parts of speech: the highly productive *-lt* suffix (*aeglaselt* ‘slowly’) and the less productive *-sti* suffix (*kiiresti* ‘quickly’). Kasik (2015, 393–395) notes that *-lt* has approx. 3500 types in a dictionary while *-sti* has approx. 250, meaning that *-lt* is 14 times more productive than *-sti*. This means the latter is more lexically and semantically restricted. Nevertheless, noone suggests that only *-lt* marks true manner adverbs, while *-sti* marks ‘oblique’ elements of a different type. Confounding productivity with quality has, however, long been the practice in argument structure domain. This has likely happened because syntactic structures are more complex than simple single-function morphemes. Their complexity means that high productivity affects them slightly differently than it does morphemes, for instance giving rise to new syntactic processes.

In any case, this thesis offers empirical support for this view that argument structures are variants of a single strategy, each characterised by a particular productivity value. We postulate that such a view is more theoretically and empirically sound than viewing them as entirely distinct phenomena.

Another takeaway from these results is that, in light of these results, regarding syntactic and semantic argumenthood as universally distinct is problematic. Languages vary greatly in terms of how the argument status of constituents is linguistically expressed, as well as the rules or tendencies for the mandatory expression of arguments. Linking syntactic obligatoriness to the quality of arguments marked by a structure, or simply using it as an indicator to detect arguments, ought to be first shown to be a suitable strategy in a language, rather than assumed to be one cross-linguistically.

### 7.2.4 Canonicity and prototypicality in gradient argument status

The present thesis offered support for the gradient nature of argument status. The predominant frameworks for discussing and expressing its gradient nature are canonical typology and prototype theory. While canonical typology and prototype theory are rather distinct, the two have been used in an analogous manner with regard to arguments. Unlike the studies conducted under these two approaches, the present study did not assess argument status based on Proto-role entailments or the morphosyntactic properties of the canonical argument structure. Hence, it is able to reflect on the outcome of the studies that did use these criteria.

Two different aspects of oblique argument structures have been regarded as non-prototypical or non-canonical — quality of argument status (Forker, 2014; Arka, 2014) and morphosyntax (Malchukov, 2005; Onishi, 2001; Lestrade, 2010). In the first approach, one refers to non-canonical arguments as weaker arguments. In the second approach, one talks about certain arguments as having non-canonical marking, making no assumptions about argument status strength.

With regard to the first approach, data presented in this thesis argues against the results of such studies (Arka, 2014; Forker, 2014), showing that the argument status of oblique arguments is as strong as that of canonical objects. The morphosyntactically based distinctions of Forker (2014); Arka (2014) did not emerge in the behavioural data of the present thesis. Interestingly, the fact that category membership degrees might not follow prototypicality was already noted by Lakoff (1987). Namely, he warned that one should not assume that greater similarity to a prototype indicates stronger membership in a category. Similarly, the results of the present thesis indicate that argument status does not appear to decrease inside the group of arguments, but only outside of it, among elements commonly viewed as adjuncts. Based on the results of the experiment in chapter 6, one can use the terms ‘non-canonical argument’ or ‘non-prototypical argument’ (section 2.4.2) to refer to various spatial adverbials which are regarded as non-arguments in traditional approaches. For instance, one may posit that participant locations are non-canonical arguments, meaning that they are weaker arguments than canonical objects, but stronger arguments than modifiers.

It is, however, rather problematic to use the label ‘non-canonical’ to describe argument status strength. Canonical typology is a framework which aims to guide and simplify typological work, offering a system of cues which may be said to indicate canonicity. This work shows that it is problematic to use morphosyntactic properties as indicators of argument status. Instead, argumenthood seems to vary based on a range of different semantic features. Many of these features are yet to be determined, this work having only scratched the surface of this field.

In any case, the labels ‘non-canonical’ and ‘non-prototypical’ are perhaps more meaningful when describing the morphosyntax of arguments rather than their argument status quality. For instance, one may wish to point out that even

though the elative argument of *sōltuma* ‘depend on’ is as strong an argument as canonical objects, the verb still has an argument structure that is less productive than the main canonical argument structure. Hence, the elative argument is non-canonically marked.

Currently, ‘non-canonical’ only reflects the (lower) syntactic productivity of the elative structure. Future studies may, however, operationalise this in a more varied manner, e.g. by quantifying semantic restrictions or by assessing the likelihood of certain types of cases occurring in argument structure (are spatial cases cross-linguistically less canonical in argument structures than, for instance comitative?). This is a productive line of research for studies interested in argument marking canonicity, because it is ill-advised to continue using the label ‘non-canonical’ to merely refer to structure productivity. One would not simply refer to less productive derivational morphemes as non-canonical and productive ones as canonical. Hence, using this label requires more elaboration. Non-canonical *in what way?*

### 7.3 The origin of obliques

Finally, the diachronic perspective on cases and argument structures links this thesis to another discussion in linguistics — a discussion on why oblique arguments have become used in language in the first place.

As has been discussed extensively in this thesis, may made two observations about argument structure distribution. First, argument structures vary in terms of productivity. Second, less productive argument structures occur more frequently with less transitive verbs than more productive structures. This means one may conceptualise two different causality relationships between them.

The first view is that oblique structures are used with less transitive verbs *because* these verbs are less transitive. Essentially this view describes something we might call a *pull effect* — namely, speakers require a variety of morphosyntactic markers (and, therefore, oblique structures), because they need to highlight unexpected subject and object properties pertaining to these verbs. Non-canonical case usage conveys non-canonicity in subjects and objects.

There is some variation inside this view. One may posit that certain verbs drawn to non-canonical structures, because they are less transitive (Tsunoda, 1981; Malchukov, 2005), or because their arguments have less Proto-role entailments (Ackerman and Moore, 2001; Lestrade, 2010). Both are regarded as variants of an entailment-based view.

The entailment-based view regards Proto-roles (Dowty, 1991) as not only describing the prototypical properties of subjects and objects, but also representing the expectations of language speakers with regard to these roles. Hence, speakers expect subjects to be prototypical, i.e. to express Agents that are sentient and volitional. They also expect objects to be prototypical, i.e. to reflect Patients that are strongly affected by the event. Upon encountering an event participant which

does not meet these expectations (e.g. a non-volitional subject), speakers tend to explicitly mark this discrepancy morphosyntactically. For instance, they might use spatial case to mark Experiencers. Hence, this view extends the framework of Dowty (1991), discussing what else might happen in addition to the expression of canonical subjects and objects.

There is some variation in terms of how these expectations are believed to work. Dowty (1991) suggested that a speaker compares constituents in a single clause in terms of Proto-Agent and Proto-Patient properties. Ackerman and Moore (2001), however, suggest that this comparison is conducted between the predicate and other predicates of the same type. Nevertheless, the important notion is that using a range of different argument structures is synchronically motivated, resulting from the communicative needs of speakers.

There are several issues with this approach. First, it expects argument structures to align with transitivity patterns. There is no quantitative data falsifying this notion (e.g. a study reviewing thousands of verb lemmas, mapping their transitivity and structures). However, there are enough examples to hypothesise that structures merely have transitivity-related tendencies, rather than being neatly mapped on transitivity degrees. Hence, many subjects with ‘unexpected’ qualities receive canonical marking, even though non-canonical marking is expected. All in all, verbs and their arguments in canonical and non-canonical argument structures are more similar to each other than this view would predict. Language-specific forces appear to affect this, meaning argument structure productivity appears to depend in other factors, rather than result from the distribution of verb lexical semantic properties.

An alternate view postulates that several different forces are at work simultaneously. The present thesis showed data which describe a *push effect*. Cases were shown have a life of their own. The functional profile of a case could affect the way its usage patterns change. Somewhat bleached functions may lead to further bleaching, perhaps resulting in the occurrence of a case in an argument structure. Further bleaching will increase the number of verbs used with that structure.

Because certain contents of a functional profile are more likely to lead to further grammaticalisation, one might regard them as pushing the case towards a lower degree of coherence and opens it up to further extension and bleaching. For instance, a semantic case may mark a bleached but still semantically coherent role in a particular type of clause (e.g. elative marking the topics of interactions, *rääki-s töö-st* talk-PST.3SG work-ELA ‘S/he talked about work’). However, the lack of spatial semantics in this clause may lead speakers to reanalyse the affix as something that is linked to the verb rather than something occurring due to its coherent semantics. In this new role, it may be assigned to new verbs, marking their participants rather than the topic of something. It is now an argument marker, now syntactically more restricted than elative in topic function, where it was also selected by nominals (*raamat sõja-st* book.NOM war-ELA ‘a book about war’).

As mentioned, instead of a pull effect serving synchronic communicative needs, this scenario describes a push effect bringing oblique argument structures into existence as results of reanalysis. Novel case functions arise for a case *because* of the nature of its previous functions and the syntactic ambiguity created by them.

In its present form, the theory of a pull effect experiences difficulties which are solved by the push effect theory. Why are some subjects and objects lacking Proto-role entailments still marked canonically? Why are non-canonical structures more predominant in some languages than others? Because argument structures vary in terms of productivity, which itself is a function of case functionality, grammaticalisation degrees, among other things. Furthermore, Nichols (1983) observes that languages vary typologically in that some may be described as ‘overlapping’ and others as ‘complementary’. For instance, the functions of structural and semantic cases overlap in Old Russian, Russian and Finnish, but the two sets of cases are functionally almost entirely complementary in Chechen-Ingush and Nanai. In these two languages, semantic cases mark only adverbials and structural cases mark grammatical relations. The pull effect cannot explain why Chechen-Ingush speakers do not need to explicitly highlight arguments referring to entities which lack the expected entailments, while Estonian speakers do. Push effect can explain it by referring to different types of case functionality and structure productivity.

However, the push effect cannot explain why upon becoming more productive, a structure is more likely to be semantically restricted to low-transitivity verbs. Until a better explanation is given, one should assume that both the push and pull effect have roles in the emergence of argument structures. For instance, it is possible that the push effect produces novel functions for cases and brings new cases in the governed structural domain. The pull effect, however, may play a role in determining the types of contexts in which this newly formed argument structure becomes used. The speed of its extension, however, is facilitated by productivity and case grammaticalisation.

## 8 SUMMARY IN ESTONIAN

### Eesti keele kohakäänded argumendistruktuuris

#### 8.1 Sissejuhatus

Käesolev doktoriväitekiri uurib eesti keele kohakäändeid ja nende esinemist funktsioonides, kus käänded ei märgi mitte asukohta (*Kalle lahkus kodust*) vaid tegusõnade kirjeldatud sündmuste osalisi ehk argumente (*Üritus sõltub ilmast*). Tegusõna argumendistruktuur on maailma keelte süntaksis keskne mõiste, sest see viitab morfoloogilisele, süntaktilisele ja semantilisele seosele, mis ühendab tegusõna umbes ühe kuni kolme fraasiga, mida võib nimetada selle tegusõna argumentideks. Enim tähelepanu saanud argumentitüübid on eesti keeles tegusõnade alused ja sihitised.

Ühelt poolt tunneb argumenti ära sellest, et sõltub tegusõnast, millises käändes argument esineb. Näiteks *sõltuma* argument esineb alati seestütlevas käändes, või *armastama* argumentid alati nimetavas ja osastavas käändes. Iga tegusõna määrab kõigest mõne üksiku fraasi käände, mitte aga kõigi osalause nimisõnafraaside käänet (*parandasin ratast toas/terrassil*). Seega tõstab tegusõna paar osalause fraasi teiste hulgas morfoloogiliselt esile.

Teiselt poolt tunneb argumenti ära selle tähenduse poolest. Argument viitab tegusõna kirjeldatud sündmuse või seisundi osalisele. Näiteks *armastama* kirjeldab seisundit, mis on võimalik ainult kahe osalise eksplitsiitsel või implitsiitsel olemasolul – peab eksisteerima keegi, kes armastab, ja keegi või miski, mida armastatakse.

Struktuuri, mille sisse on võimalik paigutada tegusõna ja kus on teatud arv kindlaksmääratud käänetes argumente, võime nimetada argumendistruktuuriks. Näiteks *vihkama* ja *armastama* esinevad sama argumendistruktuuriga, sest kodeerivad kaks identse morfoloogiaga argumenti (ühe nimetavas ja teise osastavas käändes). Tegusõnad *hoolima* ja *sõltuma* kasutavad samuti sama argumendistruktuuri, kus lisaks nimetavas käändes argumentile esineb ka seestütlevas käändes argument.

##### 8.1.1 Argumendistruktuuride süsteem

Siit kerkivad esile selle probleemi esmased piirjooned, millele käesolev doktoritöö keskendub – keeles on tavaliselt piiratud arv argumendistruktuure ehk erinevaid osalejate arvu ja käänete kombinatsioone, kuid tuhandeid erinevaid tegusõnu. Iga tegusõna kasutamine eeldab argumendistruktuuri olemasolu, seega peab iga struktuur olema kasutusel teatud hulga tegusõnadega. On põhjust eeldada, et tegu pole juhusliku jaotusega ning et eksisteerivad süsteemsed mustrid, mille järgi teatud tüüpi tegusõnu teatud struktuuridega kasutatakse, samuti ka põhjused, miks teatud käänded nendes struktuurides esinevad ja teised käänded mitte.

Tegusõnade argumendid on keeleteaduses üks prominentsem keelteüleselt uuritud süntaktiline nähtus, mistõttu on maailma keeltes omajagu tähelepanu saanud ka argumentide markeerimise varieerumine. Traditsioonilise (strukturealistliku) keeleteaduse suuna kontekstis on teadlased pigem keskendunud erinevate keelte sagedaseima (kanoonilise) argumendistruktuuri kasutusele ja varieerumisele (Blake, 2001; Ackema, 2015; Barbu and Toivonen, 2016). Eesti keeles esindavad seda nähtust struktuurid, kus alus on nimetavas või osastavas käändes alus ning sihitis omastavas, osastavas või nimetavas käändes (Metslang, 2012; Ogren, 2015; Metslang et al., 2017; Hiietam, 2003; Miljan, 2009).

Samal ajal on tähelepanu pälvinud ka asjaolu, et mitte kõik tegusõnad ei ole kasutusel kanoonilise struktuuriga, vaid kasutavad seda vaheldumisi mõne kohakäände struktuuriga (*Usun sind vs usun sinusse*), mõned tegusõnad kasutavad aga ainult kohakäände struktuuri (*sõltub ilmast*). Seda fenomeni on rohkem uuritud teiste keelte puhul (Malchukov, 2005; Lestrade, 2010; Onishi, 2001), kuid mõningast tähelepanu on see pälvinud ka eesti keeles (Klaas, 1988; Matsumura, 1994; Erelt and Metslang, 2008; Lindström and Vihman, 2017; Veismann et al., 2017).

Sellele vaatamata napib maailma keeltes sellest nähtusest süsteemseid kvantitatiivseid ülevaateid. Põhjuseks võib pidada asjaolu, et omavahel on pöördvõrdelises seoses käänete arv keeles ja keele kohta käivate lingvistiliste tööde arv. Mida laiem on keele kasutusala, seda rohkem on keeles kvaliteetseid keeleressursse (näiteks suuri morfoloogiliselt märgendatud korpusi) ning seda suurem on keelt uurivate tööde tegemise tõenäosus, kuid seda vähem on selles keeles tavaliselt ka käändeid. Tugevama keeleteadustraditsiooniga on maailmas just eelkõige väikse käänete arvuga indoeuroopa keeled. Eriti tavatsevad aga morfoloogiliselt vaesuda suured *lingua franca* tüüpi maailmakeeled (nt inglise või mandariini keel), mis paratamatult on keeleteaduslikes uurimustes eriti arvukalt esindatud. Morfoloogiline vaesumine aga tähendab, et argumente võidakse selles keeles tähistada väga väheste üksikute argumendistruktuuridega ja varieerumise uurimiseks on seega materjali kesiselt. Argumendistruktuure võidakse aga ka tähistada hoopis mittemorfoloogiliste vahendite abil (nt inglise keeles määrab sõnajärg selle, kes mida teeb: *Mary walked the dog vs The dog walked Mary*). Nähtuse enda uurimiseks on sellistel juhtudel vähem eksplitsiitset materjali, mis näitaks, milliste omaduste eristamiseks argumendistruktuuride varieerumine eksisteerib, kuidas see tekib ja kuidas seda kasutatakse.

### 8.1.2 Eesti keel kui haruldaselt kasulik allikas

Ülalmainitud põhjusel on eesti keel argumendistruktuuri uurimisel väärtuslik ressurss. Semantiliste käänete (11 käänet, mille hulka ei kuulu nimetav, osastav ja omastav) kasutamine tegusõnade argumentide markeerimiseks on eesti keele argumendistruktuurides laialdaselt levinud nähtus. Eriti tõuseb selles plaanis esile kuuest käändest koosnev kohakäänete süsteem (kolm väliskohakäänet ja kolm sisekohakäänet), mis teevad eesti keele argumendistruktuuride süsteemi ebaharili-

kult rikkalikuks ja mitmekülgseks. Samuti on põhjust eeldada, et kuigi kõik kuus kohakäänet markeerivad mõningate tegusõnade argumente, siis argumenti markeerimise funktsiooni poolest varieeruvad nad siiski tugevalt.

See rikkalik argumentide süsteem annab haruldase võimaluse heita valgust küsimustele selle kohta, kuidas ja millistel eesmärkidel maailma keeled käändesüsteeme kasutavad. Miks määravad mõned tegusõnad oma argumentidele osastava käände ja mõned seestütleva käände? Kas kohakääne saab olla täiesti ilma kohale viitava tähenduseta? Kui keelde tekib uus tegusõna, siis millistel tingimustel võivad rääkijad selle argumente tähistada kohakäändega ja milliste käänete kasutamine on sel juhul tõenäolisem? Üks oluline küsimus puudutab tegusõna ja argumenti suhet ning selle olemust. Kas argumentisuhe on kohakäänet kasutavas struktuuris nõrgem või teist laadi võrreldes sellega, kui argumenti märgib grammatiline käanne (nimetav, omastav, osastav), nagu postuleerivad peavoolukirjeldused eesti keele struktuurist (Erelt, 1989; Erelt jt. 1993; Erelt, 2003; Erelt ja Metslang, 2017; Veismann jt. 2017), ehk kas on mõtet rääkida semantilistest ja süntaktilistest argumentidest kui fundamentaalselt erinevatest nähtustest? Kas argumentisuhe on oma olemuselt binaarne või gradientne?

On tähelepanuväärne, et seniajani puuduvad ulatuslikud empiirilised argumentidstruktuuridele keskenduvad uuringud, mis püüaks neile küsimustele vastata ning avada argumenti markeerimise süsteemi sisemisi mustreid. Samuti on märkimisväärsed lüngad selles, kui palju me kõnealusest nähtusest eesti keele puhul teame. Näiteks pole empiirilisi kvantitatiivseid andmeid selle kohta, mitut ja milliseid argumentidstruktuure eesti keele rääkijad pruugivad või kui produktiivsed need struktuurid on, ehk kui levinud need on ning kui tõenäoline on nende kasutus uute verbide puhul. Me ei tea, kuidas erinevad eesti kohakäänded oma grammatiseerumistaseme poolest (ehk kui lähedale on need käänded oma arengus jõudnud grammatilistele käänetele) ning kuidas see mõjutab nende kasutust argumentidstruktuurides. Kohati on neil lünkadel ideoloogilised põhjused, sest paljud lähenemised peavad grammatiliste käänete märgitud argumente fundamentaalselt erinevaks kohakäänete märgitud argumentidest (Blake, 2001; Erelt, 1989, 2003, 2004). Teisalt on nende lünkade taga asjaolu, et keeleandmete võidukäik ja suuremat sorti korpusloome jääbki tegelikult alles viimasesse kümnendisse. Alles hiljuti on seega esmakordselt saanud võimalikuks argumentide variatiivse märkimissüsteemi (ehk argumentidstruktuuride süsteemi) süsteemne ja kvantitatiivne vaatluse alla võtmine.

## 8.2 Käesolev töö – eesmärgid ja uurimisküsimused

Doktoriväitekiri keskendub kuue eesti keele kohakäände argumentifunktsioonile. Argumentifunktsioon tähendab, et esiteks puudub kohakäändel selles kontekstis ühtne tähendus (kohatähendus või mõni muu, nt ajatähendus (*esmaspäeval ärkasin vara*)). Teiseks märgivad nad tegusõna kirjeldatud sündmuse või seisundi osalist (*hoolin isast, allun eeskirjale*).

Selliseid kohakäändes argumente on teistes keeltes nimetatud *obliikvaobjektideks* (Nichols, 1983, 1984) või *mittekanoonilisteks argumentideks* (Onishi, 2001; de Hoop ja de Swart, 2009; Lestrade, 2010). Eesti keeles on neid nimetatud *kaudobjektiks*, *kaudsihitiseks*, *indirektseks objektiks* (Kure, 1959; Saks, 1959; Klaas, 1988), aga ka *sõltuvusmääruseks* ja *rektsiooniadverbiaaliks* (Erelt, 2004; Veisermann jt. 2017). Käesolevas töös on kasutusel termin *obliikvaargument* (*oblique argument*) ja *obliikvastruktuur* (*oblique argument structure*). Rahvusvahelises perspektiivis on mõistlik vältida mõistet *kaudsihitis*, sest nii on eesti keele terminoloogia kooskõlas rahvusvaheliste terminitega, kus *indirect object* ehk *kaudsihitis* viitab üsna konkreetset tüüpi argumendile, st kolme argumendiga tegusõna Saaja argumendile (*Andsin haamri õele*).

Käesoleva töö eesmärk on võrrelda erinevate eesti keele argumendistruktuuride kasutust, et tuua välja mustrid ja tendentsid argumendistruktuuride süsteemi sisemises tööjaotuses. Et töö keskendub vaid ühele kindlale semantilisele domeenile (st käänded, mis vahel väljendavad kohatähendust) ja kohakäänete kasutusele argumendistruktuurides, siis on selles võimalik kokku siduda nii sünkrooniline kui ka diakrooniline perspektiiv. Nimelt on tõenäoline, et diakroonilised muutused väljenduvad sarnast tüüpi kääne puhul vähemalt mingil määral sarnaselt ning seega on võimalik välja tuua ühisjooni selles, kuidas nende käänete funktsioonid muutuvad.

Täpsemalt on töö jaotatud kolmeks suuremaks osaks, millest igaüks uurib argumendistruktuuride kasutust ühest konkreetsest struktuurielemendist lähtudes – nendeks on tegusõnad, käänded ja tegusõnu argumentidega ühendava kognitiivse seose (argumendistaatuse) olemus. Kolmest osast johtuvalt on doktoritööl kolm eesmärki:

1. uurida, kuidas erinevad obliikvastruktuurides ja kanoonilises struktuuris esinevad tegusõnad, keskendudes erinevatele leksikaalse semantika aspektidele ja verbide kasutusmustritele;
2. viia individuaalsete kohakäänete funktsioonid ja distributsioon korpuses kokku argumendistruktuuride levimuse ja kasutusmustritega ja näidata, kuidas ka väiksed muutused käände enda kasutuses on viinud muutusteni argumendistruktuuride süsteemis;
3. võrrelda argumendistaatuse olemust ehk tegusõna ja argumendi seose tugevust erinevates struktuurides ja demonstreerida selle sideme gradientsust.

Järgmised kolm sektsiooni arutavad töö eesmärges nii, et iga teema puhul tuuakse välja, milliste uurimisküsimuste toel eesmärki püütakse saavutada, milliste meetodite abil vastavaid küsimusi uuriti, ning millised vastused need küsimused said. Kokku moodustavad need kolm alamteemat – tegusõnad, käänded, ja argumendistaatus – ühtse pildi ühe keele argumendistruktuuride süsteemist. Kõik käesolevas töös kasutatud andmed on vabalt kättesaadavad (Aigro, 2022b,a).

## 8.3 Verbid argumendistruktuurides

Esimene osa (peatükk 4) keskendub tegusõnadele. Siin püüab väitekiri välja selgitada, kuidas erinevad kohakäänetega argumendistruktuurid (obliikvastruktuurid) ja kanoonilised struktuurid selle poolest, millised tegusõnad nendega esinevad. Peatükk puudutab nii leksikaalset semantikat kui ka tegusõnade kasutusmustreid ning keskendub kolmele uurimisküsimusele (UK):

UK1: Kas obliikvastruktuure kasutatakse transitiivsete tegusõnadega vähem kui kanoonilist struktuuri? Millisel moel on need tegusõnad vähemtransitiivsed?

UK2: Mil määral on kohale viitav tähendus obliikvastruktuuris alles?

UK3: Kas obliikvastruktuuriga tegusõnu kasutatakse teisiti kui kanoonilise struktuuriga tegusõnu? Milles need erinevused avalduvad?

UK1 osas võib püstitada hüpoteesi, et obliikvastruktuuris esinevad tegusõnad on leksikaalselt vähemtransitiivsed kui kanoonilise struktuuriga tegusõnad (Tsunoda, 1981; Malchukov, 2005; Lestrade, 2010). Transitiivsus viitab siinkohal semantilis-pragmatilisele nähtusele, mis kirjeldab nii tegusõnu, nende argumente kui ka terveid osalauseid, ning üldjoontes väljendab seda, kui võrd mõjutavad argumentide väljendatud osalised sündmusi ja kui võrd on nad ise sündmusest mõjutatud (Hopper and Thompson, 1980). Näiteks *Anni lõhkus vaasi* on väga transitiivne lause, kus on tahtlikult käituv ja tugevalt sündmust mõjutav alus, loomuliku ajalise lõpppunktiga tegevus ja sündmusest tugevalt mõjutatud sihtis (vrd nt väga madala transitiivsusega lause *Osakond allub juhatusel*).

UK2 hüpotees on, et vastupidiselt levinud arvamusele (Nichols, 1983; Blake, 2001) leidub hulk tegusõnu, mis esinevad kohakäänat kasutavates struktuurides, kuid ei väljenda mingil määral kohatähendust. Töö eeldab, et mõlemat tüüpi struktuurides (kanoonilistes omastava/osastava/nimetava sihtisega struktuurides, ja obliikvastruktuurides) võivad käänded semantiliselt märkida väga üldist argumentistaatust, mis väljendab ainult seda, et argument osaleb tegusõna kirjeldatud sündmuses, kuid ei väljenda lisanduvaid semantilisi aspekte.

UK3 on eksploratiivne uuring, kus kaardistatakse kaht tüüpi tegusõnade kasutust. Töö eeldab, et struktuuride kasutus on mõnevõrra erinev, kuid et mõned kohakäänetega struktuurid sarnanevad oma kasutuse poolest rohkem kanoonilise struktuuriga kui teised.

### 8.3.1 Meetod

UK1 ja UK2 puhul kasutati sama andmestikku, mis koosnes 104 obliikvastruktuuriga tegusõnast ja 128 kanoonilise struktuuriga tegusõnast, ning hulgast iga tegusõna kohta kodeeritud tunnustest. UK1 puhul oli relevantne tegusõnade seisundlikkus. Kirjanduse põhjal võib seisundlikkuse mõistet pidada transitiivsusega

analoogseks, sest mõlemad väljendavad skaalat, mille ühes otsas on mõjutavad ja mõjutatud argumendid, teises aga passiivsed argumendid, kus tegusõna väljendab pigem suhet või omadust, mitte aga tegevust. Erinevalt transitiivisusest, mida mõeldakse arvukate faktorite põhjal, võimaldab ühtsem ja selgem seisundlikkus tegusõnade leksikaalsest tähendusest esialgu selgemat pilti näha.

Töös kodeeriti iga verbi seisundlikkus kolmetasandilise süsteemi järgi, kus tegusõnad jaotati K-seisunditeks, D-seisunditeks ja mitteseisunditeks (Maienborn, 2008, 2019; Rothmayr, 2009). K-seisundid on omadused, mitte sündmused (*sõltuma, seisnema, omama*). D-seisundid on sündmused, mille osalised ei liigu ning mis ei ole ajaliselt piiritletud (*istuma, ootama*). Kõik muud verbid kirjeldavad mitteseisundeid (*parandama*). K-seisundid määrati nelja muutuja alusel, millest üks hõlmas vastuvõetavushinnangute katset, kus osales 43 inimest. Teised muutujad väljendasid tegusõnade semantilist tüüpi (kogejaverbid, mõõtmisverbid, omadusverbid, muud verbid), dünaamilisust (kas midagi peab tegusõna kirjeldatud sündmuses liikuma?) ja leksikaalset aspekti (kas tegusõna kirjeldatud sündmuses peab olema loomulik lõpp-punkt, vrd nt *purustama* ja *omama*). Tegusõnade hulgas, mis ei saanud K-seisundi staatust, määrati D-seisundid (mittedünaamilisi lõpp-punktita sündmusi kirjeldavad tegusõnad) ja mitteseisundid (kõik muu).

UK2 puhul kodeeriti samas tegusõnade valimis iga tegusõna semantilised rollid. Need on kasulikud, keeleteaduses laialt kasutatud mõisted, mis väljendavad erinevate lauseliikmete semantikat ühtses ja argumenditasandist täpsemas süsteemis. Tegusõna võib kodeerida Kogejat (*mulle meeldib*), Agenti (*mina lahkusin*), Saajat (*andsin emale*), Sihti (*investeerisin firmasse*), Allikat (*distantsseerusin temast*), aga ka sündmuse/seisundi osalist ehk Teemat, mis on semantiliselt “pleekinud” ehk vaesunud (*reageerisin ohule*), jne. Semantiliste rollide abil võrreldi koharollide (Siht, Allikas, Koht) ja pleekinud rollide (Teema, Patsient) esinemist kahe erineva struktuuriga esinevate tegusõnade hulgas.

Tegusõnade valim, mille põhjal käesolevas töös hinnati seisunditüüpe ja semantilisi rolle, pärineb Tasakaalukorpusest (15 mln sõna). Obliikvastruktuuri kasutatavate tegusõnade leidmiseks kaardistati iga tegusõna (mis oli sagedasem kui üks esinemine miljoni sõna kohta) ja iga kohakäände koosinemissageduse ühes osalauses. Üle 7000 tegusõnalemme kohta saadi nii kuus erinevat määra. Välja võeti tegusõnad, mis esinesid ühe kohakäändega samas osalauses rohkem kui 50% oma juhtudest (n = 200). Tegusõnade sobivus kinnitati 10 000 lause käsitsi läbi vaatamisel ja puhastamisel (täpsemalt kirjeldatud peatükis 4.1). Analoogselt saadi ka kanoonilistes struktuurides esinevate tegusõnade andmestik. Pärast erinevaid puhastusprotseduure jäi andmestikku 104 obliikvastruktuuris esinevat tegusõna ja 128 kanoonilises struktuuris esinevat tegusõna.

UK3 uuriti osalauseandmestiku põhjal, mis sisaldab neid 232 ülalkirjeldatud tegusõna. Obliikvastruktuure esindavad 4501 osalausest ning kanoonilisi tegusõnu esindavad 5364 osalausest. Mõlemad on tasakaalus valimid, kus iga tegusõna esineb umbes 50 osalauses. Hulk tunnuseid kodeeriti automaatselt (teigusõna

aeg, eitus, tegusõna ja argumendi järjekord lauses, argumendi arv, kas argument on pronoomen või ei, osalause register (ajakirjandus, ilukirjandus, teadus). Teatud tunnused kodeeriti neis 9865 lauses aga käsitsi: argumendi elusus (inimene, muu elus, abstraktne eluta ja konkreetne eluta), argumendifraasi pikkus, kaugus tegusõnast, fraasi keskse nimisõna kaugus verbist ja käände esinemise proportsioon argumendifraasi sõnades. Kõik andmed on vabalt kasutamiseks kättesaadavad (Aigro, 2022b).

### 8.3.2 Tulemused

UK1 puhul leidis kinnitust hüpotees, et kohakäänetega markeeritud argumendi-struktuurides esinevad tegusõnad väljendavad palju sagedamini seisundlikkust kui kanoonilise struktuuriga tegusõnad. Obliikvaverbide hulgas on palju sagedamini K-seisundeid (ehk omadusi väljendavaid tegusõnu) kui kanooniliste verbide hulgas. Isegi need tegusõnad, mis ei kvalifitseerunud K-seisundiks, väljendasid märkimisväärselt tihedamini vähemdünaamilisi sündmusi kui kanoonilise struktuuriga tegusõnad, mis väljendasid võrdlemisi sageli dünaamilisi sündmusi. Samas väljaspool K-seisundeid need grupid leksikaalse aspekti poolest ei erinenud. See tähendab, et kui keelde saabub uus tegusõna, siis omaduslikkus ja mittedünaamilisus tõstavad oluliselt tõenäosust, et seda hakatakse kasutama obliikvastruktuuriga.

UK2 puhul leidis samuti kinnitust hüpotees, et on olemas suur grupp tegusõnu, mis valivad kohakäändes argumendi, kuid ei väljenda mingil määral kohasemantikat. Obliikvastruktuuriga tegusõnade hulgas leidis küll neid, mis hõlmasid kohasemantikat (*osutas probleemile*), kuid sellised tegusõnad olid siinses andmestikus vähemuses. Enamik kummagi grupi tegusõnadest valis käände täielikult semantiliselt pleekinud kujul, kus selgeid tähendusgrupe pole võimalik tuvastada ja kus kääne märgib ainult argumentisuhet tegusõna ja nimisõna vahel. Seega ei saa pidada obliikvakäände algse tähenduse olemasolu läbivaks tendentsiks obliikvastruktuuride puhul, nagu seda teeb (Blake, 2001). Samuti selgus, et teistest sagedamini märgivad pleekinud rolle sellised tegusõnad, mis esinevad seestütleva käände (elatiivi) struktuuris. Enim erinevaid rolle märgib aga alaleütleva käände (allatiivi) struktuur.

UK3 puhul leiti, et obliikvastruktuurides esinevad tegusõnad on palju sagedamini kasutusel inimesi väljendavate argumentidega kui kanoonilise struktuuriga tegusõnad.<sup>37</sup> Sellest tulenevad ka mitmed teised erinevused (nt obliikvaargumentid on sagedamini ainsuses ja sõnade arvult lühemad). Lisaks leiti, et kõigi faktore löikes sarnanes elatiivi käände struktuur enim kanoonilise struktuuriga.

---

<sup>37</sup> Ka pärast seda, kui andmestikust eemaldati sageli inimestele viitavate Saaja ja Kogeja argumentidega tegusõnad (nt *meeldima*), jäi obliikvategusõnadega osalause gruppi oluliselt rohkem inimestele viitavaid argumente kui kanoonilise struktuuriga osalause gruppi. Efekt pärineb tegusõnadest nagu näiteks *andestama*, *hoolima*, *alluma*, jms, mis võimaldab postuleerida, et obliikvastruktuurid on rohkem kasutusel selleks, et rääkida inimtegevustest ja inimsuhetest.

## 8.4 Käänded argumendistruktuurides

Teine osa (peatükk 5) keskendub eesti keele kohakäänete grammatiseerumistasele. Töö uuris, kuidas kohakäänded erinevad oma grammatiseerumistasete poolest ning kuidas need erinevused mõjutavad kohakäändeid sisaldavate argumendistruktuuride kasutamist. Et ajalooliste keelematerjalide piiratud hulk ei võimalda uurida enamiku eesti keele käänete grammatiseerumist, siis diakroonilise lähenemise asemel kaardistab doktoritöö käänete grammatiseerumistaseid sünkrooniliste muutujate abil. Täpsemalt hinnati seda kolme muutuja alusel: funktsionaalne profiil, leksikaalne variatiivsus ning süntaktiline produktiivsus. Seega keskendub peatükk kolmele uurimisküsimusele (UK):

UK4: Kuidas erinevad üksteisest kohakäänete funktsionaalsed profiilid, millised kohakäänded on keelekasutuses kõige multifunktsionaalsemad, millised kõige sagedamini pleekinud?

UK5: Millistel kohakäänetel on kõige suurem leksikaalne variatiivsus (distributsioon)?

UK6: Milliste kohakäänete argumendistruktuurid on kõige produktiivsemad?

UK4 puhul kirjeldab käände funktsionaalne profiil seda, millistes funktsioonides ja kui sageli kääne mingis funktsioonis esineb. Erelt jt. (2007) toovad välja erinevad kohakäänete funktsioonid (КОИТ, АЕГ, ПӨНЮС, jne), kuid empiiriliste korpusandmeteta jääb hägusaks, millised neist on sagedased ja millised harvad funktsioonid. Funktsioonide esinemissagedused on aga grammatiseerumise kontekstis väga kõnekad. Sisuliselt kirjeldab grammatiseerumine ajapikku toimuvat funktsionaalset muutust, kus mingi element muutub grammatilisemaks. Grammatiseerumine sisaldab mitut eeldatavat muutust, sh semantilise sisu kadumine ehk pleekimine, fonoloogiline lühenemine ning süntaktiliste piirangute kasv (Heine jt. 1991). Käänded läbivad aeg-ajalt seda tüüpi muutusi ja semantilistest käänetest võivad seeläbi saada grammatilised käänded, nagu on näiteks juhtunud eesti keele osastava käändega (?Larjavaara, 1991).

Sünkrooniline funktsioonide profiil võimaldab hinnata semantiliste käänete kaugust sellel teekonnal. Mida grammatilisemaks muutuvad kohakäänded, seda väiksemal määral märgivad nad kohafunktsiooni ning rohkem märgivad nad metafoorselt veel kohatähendusega seotud, aga sellegipoolest uut tüüpi funktsioone (АЕГ, СААЈА). Edasi grammatiseerudes lahustub aga käändeliite tähendus veelgi enam ning see hakkab märkima semantiliselt täielikult pleekinud funktsioone (nt ТЕЕМА). Seega on oluline välja selgitada, millisel määral käänded juba praegu neid eri tüüpe funktsioone märgivad, sest see võimaldab hinnata kohakäänete seotuse määra argumendistruktuuridega ja seletada teisi nähtusi, mis struktuuride puhul ilmnenu on (peatükk 4).

UK5 puhul kirjeldab leksikaalne variatiivsus seda, kui paljude erinevate sõnadega käännet reaalselt kasutatakse. Kuigi erinevalt tuletusliidetest saab kääne eelduspäraselt esineda kõigi nimisõnadega, siis tegelikkuses sobivad mõned käänded teatud tüüpi sõnadega paremini kui teistega (Karlsson, 1986; Aristar, 1997). Oma algses tähenduses liituvad kohakäänded pigem elutuid konkreetseid referente väljendavate sõnadega (*laud, maja*). Kui käände funktsionaalsus laieneb ja käände tähendust tunnetatakse üldistatumalt, siis võivad need käänded esineda ka enda jaoks ebatavaliste sõnadega, näiteks inimestele viitavate sõnadega (*ema*). Sel juhul ei tõlgendata fraasi enam aga kohafraasina (*emale* ei võrdu *ema pealae peale*), vaid pigem järeldatakse, et tegu on keerulisema ja abstraktsema suhtega. See laiendab käände üldist tähendust ja ühtlasi nende lekseemide variatiivsust, millega kääne esineb (Aristar, 1997). Selle protsessi käigus kääne grammatiseerub. Järeltõlke abil saab erinevaid variatiivsusmõõdikuid kasutada selleks, et hinnata käänete suhtelisi grammatiseerumistasemeid.

UK6 puhul viitab argumentstruktuuri süntaktiline produktiivsus sellele, kui “saadaval” on see struktuur uute tegusõnade jaoks. Produktiivsuse mõistet kasutatakse sageli erinevate tuletusliidete “võimekuse” ehk uutele lekseemidele saadavaloleku hindamisel. Süntaksis võib seda aga kasutada terve struktuuri puhul (Barðdal, 2008). Mida rohkem erinevaid tegusõnu ühe struktuuriga esineb, seda produktiivsem see struktuur on ja seda tõenäolisemalt kasutatakse seda uute tegusõnadega, mis keelde näiteks laenamise teel tulevad. Struktuuri kõrgem produktiivsus viitab aga struktuuri käände kõrgemale grammatiseerumistasemele.

Eesti keele puhul on põhjust püstitada hüpotees, et seestütlev kääne ehk elatiiv (Nurka, 2014; Viht ja Habicht, 2019), alaleütlev kääne ehk allatiiv ja alalütlev kääne ehk adessiiv (Matsumura, 1994; Lindström ja Vihman, 2017) tõusevad nende kolme parameetri järgi esile kui kõige grammatiseerunud käänded.

### 8.4.1 Meetod

Esimese uuringu jaoks määrati käändefunktsioonid Tasakaalukorpusest võetud juhuvalimi põhjal. Valim sisaldab 200 lauset iga kohakäände kohta (kokku 1200 lauset) (Aigro, 2022b).

Leksikaalset variatiivsust hinnati kahe produktiivsusmõõdiku alusel – reaalne ja potentsiaalne variatiivsus. Reaalne variatiivsus kvantifitseerib erinevate käändsõnade hulga, millega käändeliidet võib leida. See mõõdab käände variatiivsust n.ö minevikus, sest see mõõdik hindab käände kasutamist tänaseni (või selle punktini, mida keeleandmed kirjeldavad).

Potentsiaalne variatiivsus aga kvantifitseerib selliste noomenite hulga, mis esinevad korpuses selles käändes vaid ühe korra (st nad on *hapax legomena* või *hapaksid*) (Baayen ja Lieber, 1991; Baayen, 1993; Baayen ja Renouf, 1996). See tähendab, et mõõdik hindab uute kontekstide hulka, milles käände kasutus on hetkel harv, kuid võib tulevikus sageda. Mõõdik hindab seega käände kasutuse tulevikusuundi.

Süntaktilist produktiivsust hinnati juhuslikus tegusõnalemme valimis. Ta-sakaalukorpusest võeti välja tegusõnad, mis esinesid rohkem kui ühe korra miljoni sõna kohta ( $n = 3043$ ). Neist võeti omakorda 1000 verbilemmaga juhuvalim. Selles valimis kodeeriti käsitsi iga tegusõna argumentide arv ja nende käanded, ehk siis verbi argumentidstruktuur (Aigro, 2022b).

## 8.4.2 Tulemused

UK4 sai vastuseks, et elatiiv ja allatiiv ilmuvad kohatähenduses oluliselt harvem ja pleekinud tähenduses oluliselt sagedamini kui teised kohakäanded. Umbes pooltel oma esinemisjuhtudest puudub neil käännetel igasugune kohasemantika. Siia alla loeti aga ka semantiliselt ühtne elatiivi funktsioon, kus elatiiv tähistab millegi teemat (*lugesin roomlastest*). Täielikult pleekinud funktsioonis esineb elatiiv kolmandikul oma juhtudest ja allatiiv pisut vähem kui pooltel oma juhtudest. Sel juhul märgivad need käanded vaid süntaktilist sidet tegusõna ja noomeni vahel. Lisaks leiti, et elatiiv on kasutusel palju suurema hulga erinevate funktsioonidega ( $n = 10$ ) kui teised käanded (5–8). Funktsionaalsete profiilide järgi on elatiiv ja allatiiv grammatiseerumise poolest pigem semantiliste ja grammatiliste käännete vahel, mitte aga pelgalt kohakäanded.

UK5 sai vastuseks, et elatiiv ja allatiiv on leksikaalselt oluliselt variatiivsemad kui teised kohakäanded. Elatiiv ületab selles plaanis isegi partitiivi, mis on eesti keeles kõige sagedasem objektikäane ning mille kõrge grammatiseerumisastme järgi võiks eeldada, et kehtib vastupidine olukord.

UK6 puhul leiti aga, et eesti keeles on kasutusel vähemalt 18 erinevat argumentidstruktuuri, millest kaks (kanoonilist) esinevad 88%-ga verbidest. Üks neist kodeerib ühe (nimetavas käändes) osalise ja teine kaks osalist (üks nimetavas või osastavas, teine omastavas, nimetavas või osastavas, sealjuures ei eristatud siin uurimuses partitiivitegusõnu sihitise käände eristavat vaheldumist lubavatest tegusõnadest). See tähendab, et 12% eesti verbidest on kasutusel obliikvastruktuuriga. Kolm kõige produktiivsemat obliikvastruktuuri kodeerivad komitatiivi, elatiivi ja allatiivi.

Kokkuvõttes leidis töö, et elatiiv ja allatiiv on kõige grammatiseerunud käanded. Seda võib järeldada asjaolust, et kõigi kolme siin peatükis uuritud muutuja lõikes tõusid esile just need kaks käanet. Samuti leiti eelmise peatüki (peatükk 4) tegusõna leksikaalse semantika uuringutes, et need kaks käanet ilmuvad argumentidstruktuuris kõige sagedamini ilma kohale viitava tähenduseta, mis on tõenäoliselt grammatiseerumise tulemus. Antud tulemus selgitab ka 4. peatüki tegusõnade kasutusmuustrite uuringu tulemusi, kus leiti, et elatiivi kasutava argumentidstruktuuriga tegusõnade kasutus on kõige sarnasem kanoonilise struktuuriga tegusõnade kasutusega.

## 8.5 Argumendistaatus argumendistruktuurides

Käesoleva väitekirja kolmas suur uurimisteema on argumendistaatuse olemus. Üldise arusaama järgi on argumendistaatus binaarne. See tähendab, et miski kas on või ei ole tegusõna argument. Näiteks lauses *vaatasime emaga filmi* on *filmi* tegusõna argument ja *emaga* ei ole. Lisaks ei loeta traditsioonilises käsitluses kohakäändes esinevat argumenti päris sama tüüpi argumendiks kui näiteks osastavas käändes kanoonilist sihitist. Näiteks lausetes *kaardistab asjaolu* ja *seisneb asjaolus* on sõna *asjaolu* (süntaktiline) argument ainult esimeses lauses, kuid teises võib seda vaid pidada semantiliselt argumendiks, sj need kaks kategooriat arvatakse kodeerivat erinevat tüüpi või erineval määral argumendistaatust (Erelt, 2004; Veismann jt. 2017).

Seda eristust põhjendatakse sageli erinevate nähtustega, mille poolest kanooniline struktuur obliikvastruktuuridest erineb. Esiteks esinevad kanoonilises struktuuris palju grammatilisemad käänded (nimetav, osastav, omastav), mis moodustavad ühtse variatsioonisüsteemi ehk sihitise käänet eristava markeerimissüsteemi (*sööb leiba, sõi leiva, söö leib!*) (Klaas, 1988). Teiseks on kanooniline struktuur hõlmatud mitmetesse süntaktilistesse protsessidesse, millega obliikvastruktuurid seotud ei ole, seahulgas näiteks verbi ühildumine subjektiga arvus, objekti muutumine nimetavakäändeliseks aluseks passiivses konstruktsioonis, ja omastavakäändeline alus nominaliseeritud tegusõna puhul (Erelt, 1989, 2003, 2004). Nii siis ennustab see lähenemine, et kanoonilises struktuuris on argumendisid kas tugevam või hoopis teist laadi kui obliikvastruktuuris ja see lähenemine esindab ka eesti keeleteaduse peavooluvaadet.

Alternatiivne vaade püstitab aga hüpoteesi, mille järgi on nii kanoonilises kui ka obliikvastruktuuris samalaadne ja sama tugev argumendisid (Nichols, 1983; Kure, 1959). Erinevused süntaktilises käitumises (see, et obliikvastruktuurid ja kanooniline struktuur pole hõlmatud samadesse protsessidesse) on sel juhul seletatavad sellega, et argumendistruktuurid erinevad produktiivsuse poolest (nagu demonstreeriti 5. peatüki viimases uuringus). Kui üks struktuur on keeles palju prominentsem, siis võibki eeldada, et seda on võimalik kasutada erinevamate verbidega kui vähem kasutatud struktuure, ning et kõige sagedamini kasutatava struktuurina on see ka suuremal määral integreeritud süntaktilistesse protsessidesse. Selle vaate järgi ei illustreeri süntaktiline asümmeetria struktuuride vahel mitte erinevat tüüpi või erineva tugevusega argumendisid, vaid see on hoopis erinevate produktiivsustasemetete tulemus.

Selle vaate järgi on argumendistruktuur laias laastus samalaadne lingvistiline tööriist nagu näiteks tuletusliited. Erinevad produktiivsustasemed on derivatsioonilises morfoloogias igapäevane nähtus. Näiteks tekib pea kõikides inimkeeltes vajadus väljendada viisimäärusi ja sageli on selle jaoks grammatiseerunud käepärased vahendid ehk erinevad viisimäärusliited. Eesti keeles on näiteks olemas kaks liidet, mida saab erinevat tüüpi sõnadele liita, et selliseid määrusi tuletada: äärmiselt produktiivne *lt*-liide (aeglaselt) ja palju vähem produktiivne *sti*-liide

(*kiiresti*). Esimene on 14 korda produktiivsem kui teine (Kasik, 2015, 393–395). Keeleteadlased ei väidaks, et *-sti* ei moodusta “tõelisi” viisimäärusi ja et *kiiresti* on mingis mõttes teist laadi või nõrgem viisimäärus kui *aeglaselt*, hoolimata sellest, et *sti*-liidet pole võimalik kasutada nii paljudes erinevates kontekstides kui *lt*-liidet, ka on *-sti* semantiliselt piiratum. Tegelikult mõistetakse, et mõlemad liited moodustavad sama tüüpi elemente (viisimäärusi), kuid teevad seda erineva võimsusega. Seda võimsust on meil võimalik mõõta ja kvantifitseerida produktiivsusemõõdikute alusel. Samuti on võimalik uurida, kuidas ja millistes valdkondades produktiivsusefektid keeles avalduvad. Nagu siin töös aga läbivalt seletatud, siis väiksem produktiivsus toobki automaatselt kaasa erinevad leksikaalsed ja süntaktilised piirangud.

Neid kaht vaadet, millest üks näeb kanoonilisi argumente obliikvaargumentidest erinevana ja teine näeb neid samaväärsena, on võimalik katseliste vahenditega testida. Nagu illustreerivad arvukad katsetel põhinevad uuringud (Clifton jt. 1991; Schütze ja Gibson, 1999; Koenig jt. 2003; Conklik, 2004; Grodner, 2005), siis argumendiseost on võimalik hinnata ka muude tunnuste abil kui verbide semantika, valitud käänded või morfosüntaks. Et tarvis on omavahel võrrelda suhteliselt erinevaid struktuure, mida oleks keeruline reaajas protsessimist uuriva (*online*) eksperimendi jaoks vajalikul hulgal identses kontekstis esitada, siis kasutab käesolev väitekiri *offline* eksperimentaalset meetodit ehk argumendihinnanguid. Selliste meetodite abil on näiteks näidatud, et mõnede tegusõnade puhul on Instrumendid tugevamad argumendid kui teiste puhul (Koenig jt. 2003; Rissman jt. 2015), aga ka et teatud tüüpi asukohad võivad olla poolargumendid (Koenig jt. 2003).

Teine eesmärk käesoleva peatüki eksperimendil on näidata argumendistaatuse gradientsust. Selle jaoks hõlmab katse mitmeid eri tüüpi kohakäändes elemente ning püüab näidata süsteemseid erinevusi selles, kui võrd peetakse neid lause osalisteks. Ennustame, et mõnda tüüpi elemente võibki pidada poolargumentideks. Meetodist, uurimisküsimustest ja hüpoteesidest on juttu järgmises alapeatükis.

### 8.5.1 Meetod ja küsimused

Eksperiment, millega hüpoteesi testiti, hõlmas seda, et katseosalised andsid hinnanguid sellele, kas nende meelest oli ekraanil ilmuvas lauses oleva fraasi näol tegu tegusõna kirjeldatud sündmuse või seisundi osalisega või mitte. Kuigi katsejuhised palusid hinnata, kas fraasid on “argumendid”, siis argumentsus oli juhistes täielikult defineeritud läbi osalemise kontseptsiooni.

Tartu Ülikooli foneetikalaboris toimunud katses osales 42 inimest. Enne katse algust said nad juhised, olid analoogsed Rissman jt. (2015) juhistega. Argumente tutvustati neljast küljest: 1) argumendid kirjeldavad osalisi, ilma milleta ei saaks tegusõna kirjeldatud sündmus/seisund toimuda, 2) argumendid võivad ilmuda tegusõnaga samas lauses, kuid neid võib ka välja jätta, 3) mõned tegusõnad võtavad

ainult ühe argumendi, mõned aga kaks või kolm, 4) viisimäärused või põhjust väljendavad fraasid pole tegusõna jooksma argumendid, aga see ei tähenda, et nad pole kunagi ühegi tegusõna argumendid.

Iga inimene nägi 176 eksperimentaallauset ja 90 täitelauset. Eksperimentaallausetes ilmus nelja tüüpi tegusõnu: 1) kohatähenduseta obliikvategusõnad (*al-luma, hoolima*), 2) abstraktse kohatähendusega obliikvategusõnad (*osutama, distantseeruma*), 3) tegusõnad, mis ilmuvad koos sellise asukohafrasiga, mis väljendab ühe osalise suunda, ja 4) tegusõnad, mis ilmuvad koos sellise asukohafrasiga, mis ei väljenda ühe osalise suunda. Viimased kaks tegusõnatüüpi ilmuvad asukohafrasidega, mis kirjeldavad kahest osalisest vaid üht. Ühe osalise suunaga asukohad on näiteks fraasides *riputab seinale* ja *trükkib paberile*, kus asukoht ei kirjelda tegusõna alust (trükkija ise ei asu paberil) ja sellel on suunasemantika. Ühe osalise suunata asukohad on näiteks fraasides *peseb kausis* ja *keedab potis*, kus asukohad kirjeldavad samuti vaid sihitist, aga suunaelement puudub. Täitelausetes esinesid kanooniliste sihitistega tegusõnad ja nende mõte oli mitmekesistada üldist lausegruppi ja ka moodustada kontrollgrupp, sest sihitise argumendiks valimise määr peaks olema võrdlemisi kõrge. Uurimisküsimusi oli kaks:

UK7: Kas abstraktne kohatähendus viitab mõnede obliikvategusõnade puhul nõrgemale argumendistaatusele?

UK8: Kas kohatähenduse olemasolu viitab mõnedes asukohafrasides tugevemale argumendistaatusele?

Eeldati, et kohatähenduseta obliikvaargumendid on sama tugevad argumendid (ehk valitud argumendiks sama hulga katses osalejate poolt) kui kanoonilised sihitised. Ennustati aga, et metafoorne kohatähendus teises tegusõnagrupis alandab nende argumendiks valimise määra (UK7). Eeldatakse ka, et suunasemantika olemasolu ühes asukohagrupis mõjutab nende argumendistaatuse tugevust ja et neid valitakse sagedamini kui suunasemantikata asukohti (UK8).

## 8.5.2 Tulemused

Esiteks kinnitas katse, et argumendiskaala äärmustes on samalaadsed elemendid kui lingvistiline traditsioon võiks eeldada, ehk kanoonilised sihitised ja kanoonilised adverbiaalid (ajafraasid, viisimäärused, jms). Teiseks näitas katse, et kohatähenduseta obliikvaargumente valiti argumendiks sama sagedalt kui kanoonilisi sihitisi, ehk umbes 90 protsendil juhtudest.

Esimene hüpotees ei leidnud aga kinnitust. Ennustati, et abstraktse kohatähendusega obliikvaargumendid (*osutab probleemile*) on nõrgemad argumendid kui kohatähenduseta obliikvaargumendid (*sõltub ilmast*). See ei osutunud tõeks, mõlemad valiti argumendiks umbes sama sagedalt (90% juhtudest), ehk sama sagedalt kui sihitised (90%). See on üllatav tulemus, sest see viitab, et abstraktne kohatähendus ei ole rääkija jaoks kuigi erinev täiesti pleekinud tähendusest.

Nagu uuringust selgus, siis on konkreetne kohatähendus obliikvaargumentidest väga erinev. Mõlemad konkreetse kohatähendusega grupid (suunaga osaleja asukohad ja suunata osaleja asukohad) hinnati argumentdiks harvem kui obliikvaargumentid. Samuti leidis kinnitust teine hüpotees – suunasemantika olemasolu viis tugevama argumentdistaatuseni. See tähendab, et suunaga asukohti (*peidab taskusse*) hinnati argumentdiks rohkemate osalejate poolt (64%) kui suunata asukohti (*keedab potis*, 39%). Kõik muud uuringus esinenud adverbiaalitüübid hinnati valdavalt mitteargumentideks, enamjaolt valis neid argumentdiks vaid paar protsenti osalejatest.

See uuring lisandub olemasolevatele uuringutele (Koenig jt. 2003; Rissman jt. 2015), mis annavad põhjust mõelda argumentdistaatusest kui loomu poolest gradientsest nähtusest.

## 8.6 Kokkuvõte

Väitekiri seadis kolm suurt eesmärki, mille raames esitati kaheksa uurimisküsimust. Kaheksale uurimisküsimusele vastati seitsme erineva uuringuga, millest viis on korpusuuringud, üks on eksperiment ja üks sisaldab nii korpusmaterjali kui eksperimenti. Korpusuuringute aluseks on omakorda viis erinevat korpusandmestikku. Käesoleva töö andmestikud on vabalt kasutamiseks saadaval (Aigro, 2022b,a).

Doktoritöö esimene eesmärk oli uurida, kuidas obliikvategusõnad ja kanoonilise struktuuriga tegusõnad leksikaalse semantika ja kasutuse poolest erinevad. Selle raames vastati kolmele uurimisküsimusele, millest selgusid järgnevad asjaolud:

- tegusõna saab tõenäolisemalt kohakäänedega argumentstruktuuri siis, kui tegusõna väljendab äärmiselt staatilist seisundit (omadust), või kui see väljendab mittedünaamilisi sündmusi;
- hulk tegusõnu ilmub kohakäänedega struktuuridega, kuid ei väljenda kohasemantikast – elatiivi struktuuriga on see kõige prominentsem;
- obliikvastruktuuris esinevate tegusõnade argumentide hulgas on suurem hulk inimesi tähistavaid sõnu kui kanooniliste struktuuridega esinevate tegusõnade argumentide hulgas, ja elatiiviga struktuuri kasutus sarnaneb enim kanoonilise struktuuri kasutusele.

Teine eesmärk oli ühendada kohakäänete individuaalne diakrooniline perspektiiv nende esinemise argumentstruktuuride süsteemis. Selle raames vastati kolmele uurimisküsimusele ning jõuti järgmistele järeldustele:

- elatiivi ja allatiivi funktsionaalsed profiilid sisaldavad kõrgeimat määra pleekinud funktsioone, aga ka kohatähenduseta funktsioone, ning elatiivil on kõigest kohakäänetest kõige laiem funktsionaalne profiil;

- elatiivil ja allatiivil on kohakäänetest kõige suurem leksikaalne variatiivsus;
- elatiivi ja allatiivi struktuurid on teiste kohakäänete struktuuridest produktiivsemad.

Kolmas eesmärk oli uurida obliikvaargumentide argumentistaatust ning näidata kohakäänete erinevate funktsioonide perspektiivis seda, millised elemendid paigutuvad argumentiskaala erinevatesse osadesse (st mis on tugevad argumentid, mis pooltugevad argumentid ja mis ei ole argumentid). Nende punktide uurimiseks püstitati kaks uurimisküsimust, millele vastamiseks korraldati eksperiment. Selle käigus selgusid järgmised punktid:

- tulemuste põhjal võib argumentistaatust pidada loomu poolest gradientseks nähtuseks;
- kohatähenduseta obliikvaargumentid on sama tugevad argumentid kui kaanonilises käändes sihitised ning sama kehtib abstraktse kohatähendusega obliikvaargumentide kohta;
- asukohta väljendavates fraasides tugevdab suunasemantika nende argumentistaatust, kuid üldiselt on kõik füüsilist asukohta väljendavad fraasid nõrgemad argumentid kui obliikvaargumentid;
- erinevad muud määrused, nt aega, põhjust, sündmuse asukohta kirjeldavad fraasid, on mitteargumentid.

Kokkuvõttes esitab töö ülevaatliku pildi sellest, kuidas rikkalikust käändesüsteemist võib tekkida rikkalik argumentistruktuuride süsteem. Töö käib üksipulgi läbi käänete teekonna kohatähenduse dimensioonis, kus kohakäänded on erineval määral saanud endale grammatilisi funktsioone. Töö näitab, millistele tasemetele on erinevad käänded jõudnud ja millised faasid on selle jaoks tulnud läbida. Töö tulemused on ka mitmel viisil kõnekad grammatiseerumisraamistikele. Ühelt poolt pakub see välja viisi käändegrammatiseerumise operatsionaliseerimiseks sünkroonilistes keeleandmetes, teiselt poolt aga kirjeldab käänete grammatiseerumist sarnaselt Nicholsi ja Timberlake'i (1991) uuringule. Nimelt ei pruugi käänded läbida teleoloogilist muutust grammatilise seisuse suunas, vaid tegelikult on tegu käändefunktsioonide laienemisega. See tähendab, et eesti keeles eksisteerivad multifunktsionaalsed käänded, mis ei ole loobunud kohafunktsioonist, kui nad grammatilisi funktsioone juurde on saanud. Kuigi grammatikad tavatsevad käändeid lahterdada grammatilisteks ja semantilisteks, siis elatiivi ja allatiivi näol on tegu segakäänetega. Tulevaste uurimuste perspektiivis tähendab see, et enne käändejaotuse eeldamist peaks uuringud küsima küsimuse, kas vaadeldavas keeles on käänded monofunktsionaalsed või multifunktsionaalsed. Selle pakkus tüpoloogiliseks tunnuseks välja juba Nichols (1983, 1984).

Samuti vaatab töö detailselt seda grammatilist faasi, kuhu mõned käanded on jõudnud vähemal määral ja teised rohkemal määral. Töö näitab kohakäändestruktuuride üldist piiratust tegusõnade leksikaalse semantika osas ning näitab kvantitatiivsete andmete toel elususe rolli kohakäänete grammatiseerumisel. Aristar (1997) kirjeldas kohakäände kombineerumist inimesi tähistavate sõnadega kui indikatsiooni, et kääne grammatiseerub. Käesolev töö näitas, et obliikvastruktuurides tegusõnad on tõenäolisemad esinema inimesi väljendavate argumentidega, isegi kui jätta välja Saaja ja Kogeja argumente kodeerivad tegusõnad. Seega võib arvata, et vähealt osaliselt juhib sama tendents grammatiseerumisprotsessi ka eesti keeles.

Tööst kerkib ka perspektiiv, milles erinevad argumendistruktuurid on samaväärsed morfosüntaktilised tööriistad, mis võimaldavad rääkijatel kasutada tegusõnu ökonoomsel moel. Suure hulga tegusõnade jaoks on olemas vaid loetud hulk argumendistruktuure ning need struktuurid varieeruvad oma produktiivsuse poolest. Produktiivsuserinevuste üheks tagajärjeks on süntaktiline asümmeetria struktuuride vahel – mõned struktuurid on seotud teiste süntaktiliste protsessidega (nt passiviseerumine), teised aga pole. Kõik struktuurid, sh kanoonilised ja obliikvastruktuurid, väljendavad aga sisuliselt sama tüüpi argumendisuhet.

## REFERENCES

- Aarts, Bas. 2007. *Syntactic gradience*. Oxford University Press, Oxford.
- Ackema, Peter. 2015. Arguments and adjuncts. In Tibor Kiss and Artemis Alexiadou, editors, *Syntax – Theory and Analysis: An International Handbook, Vol 1*, De Gruyter Mouton, Berlin, 246–274.
- Ackema, Peter and Ad Neeleman. 2004. *Beyond morphology: Interface conditions on word formation*. Oxford University Press, Oxford.
- Ackerman, F. and J. Moore. 1999. Telic Entity. In Miriam Butt and Tracy Holloway King, editors, *Proceedings of the LFG99 Conference*. CSLI Publications, Stanford.
- Ackerman, F and J Moore. 2001. *Proto-Properties and Grammatical Encoding: A Correspondence Theory of Argument Selection*. CSLI Publications, Stanford.
- Ágel, Vilmos. 2000. *Valenztheorie*. Narr, Tübingen.
- Aigro, Mari. 2018. Kvantitatiivne pööre: vana tüli uues kuues. *Keel ja Kirjandus* 8-9: 747–764.
- Aigro, Mari. 2020. Polar question particles and their sources: A semantic approach to grammaticalisation. *Transactions of the Philological Society* 118(2): 237–253.
- Aigro, Mari. 2022a. Argument status of oblique arguments (experiment datasets).
- Aigro, Mari. 2022b. Estonian cases as argument markers (datasets of corpus studies).
- Aigro, Mari and Virve-Anneli Vihman. 2022. Overabundance in Estonian noun paradigms (in review). *Word Morphology*.
- Aigro, Mari and Virve-Anneli Vihman. 2023. Oblique complements in Estonian: A corpus perspective (accepted). *Journal of Linguistics*.
- Aissen, J. 1999. Markedness and Subject Choice in Optimality Theory. *Natural Language and Linguistic Theory* 17: 673–711.
- Altmann, Gerry and Mark Steedman. 1988. Interaction with context during human sentence processing. *Cognition* 30: 191–238.
- Aristar, Anthony Rodrigues. 1997. Marking and Hierarchy Types and the Grammaticalization of Case-Markers. *Studies in Language* 21(2): 313–368.
- Arka, I Wayan. 2014. Locatives and the argument-adjunct distinction in Balinese. *Linguistic Discovery* 12(2): 56–84.
- Baayen, Harald. 1993. On frequency, transparency and productivity. In Geert E. Booji and Jaap van Marle, editors, *Yearbook of Morphology 1992*, Kluwer Academic Publishers, Dordrecht, 181–208.
- Baayen, Harald and Rochelle Lieber. 1991. Productivity and English derivation: A corpus-based study. *Linguistics* 29(5): 801–844.
- Baayen, R Harald and Antoinette Renouf. 1996. Chronicling the Times : Productive Lexical Innovations in an English Newspaper. *Language* 72(1): 69–96.
- Barbu, Roxana-Maria and Ida Toivonen. 2016. Event participants and linguistic arguments. *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (May): 1961–1966.
- Barðdal, Johanna. 2006. Predicting the productivity of argument structure constructions. In Zhenya Antić, Charles B. Chang, Emily Cibelli, Jisup Hong, Michael J. Houser, Clare S. Sandy, Maziar Toosarvandani, and Yao Yao, editors, *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*. BLS and LSA, Berkeley, volume 32, 467–478.
- Barðdal, Johanna. 2008. *Productivity. Evidence from case and argument structure in Icelandic*. John Benjamins, Amsterdam and Philadelphia.

- Bates, D. and M Maechler. 2009. lme4: linear mixed effects models using S4 classes.
- Bauer, Laurie. 2001. *Morphological productivity*. Cambridge University Press, Cambridge.
- Belletti, A and L Rizzi. 1988. Psych verbs and theta theory. *Natural Language and Linguistic Theory* 6(3): 291–352.
- Blake, Barry J. 2001. *Case*. Cambridge University Press, Cambridge.
- Blansitt, Edward. 1988. Datives and allatives. In Michael Hammon, Edith A. Moravcsik, and Jessica Wirth, editors, *Studies in Syntactic Typology*, John Benjamins Publishing Company, Amsterdam and Philadelphia, 173–191.
- Boland, Julie E. 2005. Visual Arguments. *Cognition* 95(3): 237–274.
- Boland, Julie E. and Heather Boehm-Jernigan. 1998. Lexical Constraints and Prepositional Phrase Attachment. *Journal of Memory and Language* 39(4): 684–719.
- Bolinger, Dwight. 1961. *Generality, gradience, and the all-or-none*. Mouton, The Hague.
- Bouchard, Denis. 1995. *The semantics of syntax: A minimalist approach to grammar*. University of Chicago Press, Chicago.
- Bremel, Neil and Luděk Knittl. 2012. Morphosyntactic variation and syntactic constructions in Czech nominal declension: corpus frequency and native-speaker judgments. *Russ Linguist* 36: 91–119.
- Brown, Dunstan, Marina Chumakina, and Greville G. Corbett. 2013. Canonical Morphology and Syntax. *Canonical Morphology and Syntax* 1–336.
- Butt, Miriam. 2006. *Theories of Case*. Cambridge University Press, Cambridge.
- Butt, Miriam. 2008. Case in lexical-functional grammar. In Andrej Malchukov and Andrew Spencer, editors, *The Oxford Handbook of Case2*, Oxford University Press, Oxford, 59–71.
- Bybee, Joan. 1985. *Morphology: A study of the relation between meaning and form*. John Benjamins, Amsterdam and Philadelphia.
- Bybee, Joan L. and Carol Lynn Moder. 1983. Morphological classes as natural categories. *Language* 59(2): 251–270.
- Cann, Ronnie and Merilin Miljan. 2012. Differential case-marking: Syntactic descriptions and pragmatic explanations. *Linguistic Review* 29(4): 585–605.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. MIT Press, Cambridge.
- Chomsky, Noam. 1981. *Lectures on Government and Binding. The Pisa lectures..* Foris Publications, Dordrecht, 5th editio edition.
- Chomsky, Noam. 1995. *The Minimalist Program*. MIT Press, Cambridge.
- Chomsky, Noam. 2002. *On nature and language*. Cambridge University Press, Cambridge.
- Clifton, Charles, Shari Speer, and Steven P. Abney. 1991. Parsing arguments: Phrase structure and argument structure as determinants of initial parsing decisions. *Journal of Memory and Language* 30(2): 251–271.
- Condoravdi, Cleo and Jean Mark Gawron. 1996. The context-dependency of implicit arguments. In Makoto Kanazawa, Christopher Pinon, and Henriette de Swart, editors, *Quantifiers, deduction and context*, CSLI Publications, Stanford.
- Conklin, Kathy, Jean Pierre Koenig, and Gail Mauner. 2004. The role of specificity in the lexical encoding of participants. *Brain and Language* 90(1-3): 221–230.
- Corbett, Greville G. 2005. The canonical approach in typology. In Zygmunt Frajzyngier, Adam Hodges, and David S Rood, editors, *Linguistic diversity and language theories*, John Benjamins Publishing Company, Amsterdam and Philadelphia, 72, page 430.
- Corbett, Greville G. 2006. *Agreement*. Cambridge University Press, Cambridge.

- Corbett, Greville G. 2007. Canonical typology, suppletion, and possible words. *Language* 83(1): 8–42.
- Corver, Norbert and Henk van Riemsdijk, editors. 2001. *Semi-lexical categories: The function of content words and the content of function words*. Mouton de Gruyter, Berlin.
- Croft, William. 1991. *Syntactic categories and grammatical relations. The cognitive organization of information*. University of Chicago Press, Chicago and London.
- Croft, William. 1998. Event structure in argument linking. In M. Butt and W. Geuder, editors, *The projection of arguments*, CSLI Publications, Stanford, 21–65.
- Croft, William. 2001. *Radical construction grammar*. Oxford University Press, Oxford.
- Cruse, D. A. 1973. Some Thoughts on Agentivity. *Journal of Linguistics* 9: 11–23.
- Curme, G. O. 1935. *A grammar of the English language: Parts of speech and accidence*. D. C. Heath, Boston.
- Danesh, F. 1966. The relation of centre and periphery as a language universal. *Travaux linguistiques de Prague* 2: 9–21.
- Davidson, Donald. 1967. The logical form of action sentences. In Nicholas Rescher, editor, *The logic of decision and action*, University of Pittsburgh Press, Pittsburgh, 81–95.
- Davis, A.R. 2001. *Linking by Types in the Hierarchical Lexicon*. CSLI Publications, Stanford.
- Davis, A.R. and Jean-Pierre Koenig. 2000. Linking as Constraints on Word Classes in a Hierarchical Lexicon. *Language* 76: 56–91.
- de Hoop, Helen and Peter de Swart. 2009. Cross-linguistic variation in Differential Subject Marking. In Helen de Hoop and Peter de Swart, editors, *Differential subject marking*, Springer, New York, 1–16.
- de Hoop, Helen and Andrej Malchukov. 2007. On fluid differential case marking: A bidirectional OT approach. *Lingua* 117(9): 1636–1656.
- DeArmond, Richard C. and Nancy Hedberg. 1998. On Complements and Adjuncts. In *Proceedings of the Canadian Linguistic Association Annual Conference, Cahiers Linguistiques d'Ottawa*. University of Ottawa, Ottawa.
- Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3): 547–619.
- Dowty, David. 2000. The dual analysis of adjuncts/complements in Categorical Grammar. *ZAS Papers in Linguistics* 17: 53–78.
- Eckhoff, Hanne M. and Laura A. Janda. 2014. Grammatical profiles and aspect in Old Church Slavonic. *Transactions of the Philological Society* 112(2): 231–258.
- Erelt, Mati. 1989. *Eesti lauseliikmeister (Grammatical Relations in Estonian)*. Eesti NSV Teaduste Akadeemia, Tallinn.
- Erelt, Mati. 2003. Süntaks (Syntax). *Emakeele Seltsi Aastaraamat* 48: 63–97.
- Erelt, Mati. 2004. Lauseliigendusprobleeme eesti grammatikas (Issues with Parts of Speech in Estonian Grammar). In Liina Lindström, editor, *Lauseliikmeister eesti keeles*, Tartu University Press, Tartu, 7–16.
- Erelt, Mati. 2017a. Ellipsis. In *Eesti keele süntaks*, Tartu University Press, Tartu, 590–601.
- Erelt, Mati. 2017b. Õeldis. In Mati Erelt and Helle Metslang, editors, *Süntaks*, Tartu University Press, Tartu, 93–239.
- Erelt, Mati. 2017c. Sissejuhatus süntaksisse. In *Eesti Keele Süntaks (Eesti keele varamu 3.)*, Tartu University Press, Tartu, 53–92.
- Erelt, Mati, Tiiu Erelt, and Kristiina Ross. 2007. Eesti keele käsiraamat.
- Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja

- Tael, and Silvi Vare. 1993. *Eesti keele grammatika II*. Eesti Teaduste Akadeemia, Tallinn.
- Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael, and Silvi Vare. 1995. *Eesti keele grammatika I*. Eesti Teaduste Akadeemia, Tallinn.
- Erelt, Mati and Helle Metslang. 2008. Koguja Vormistamine Eesti Keeles. In Mati Erelt, editor, *Emakeele Seltsi Aastaraamat 54*, Teaduste Akadeemia Kirjastus, Tallinn.
- Erelt, Mati and Helle Metslang, editors. 2017. *Eesti Keele Süntaks (Eesti Keele Varamu 3)*. Tartu University Press, Tartu.
- Erelt, Mati, Helle Metslang, and Helen Plado. 2017. Alus. In Mati Erelt and Helle Metslang, editors, *Eesti Keele Süntaks (Eesti keele varamu 3.)*, Tartu University Press, Tartu, 240–252.
- Ernestus, Mirjam, Harald Baayen, and Rob Schreuder. 2002. The recognition of reduced word forms. *Brain and Language* 81(1-3): 162–173.
- Evans, Nick. 1997. Role or cast. In A. Alsina, J. Bresnan, and P. Sells, editors, *Complex Predicates*, CSLI Publications, 397–430.
- Fábregas, Antonio and Rafael Marín. 2012. State nouns are Kimian states. In Irene Franco, Sara Lusini, and Andres Saab, editors, *Romance Languages and Linguistic Theory 2010: Selected papers from 'Going Romance' Leiden 2010*, John Benjamins, Amsterdam and Philadelphia, 41–64.
- Fillmore, Charles J. 1968. The Case for Case. In E Bach and R. T. Harms, editors, *Universals in linguistic theory*, Holt, Rinehart and Winston, New York, 1–88.
- Fillmore, Charles J. 1969. Types of lexical information. In Ferenc Kiefer, editor, *Studies in syntax and semantics*, Reidel, Dordrecht, 109–137.
- Fillmore, Charles J. 1977. The Case for Case Reopened. In Peter Cole and Jerrold M. Sadock, editors, *Grammatical Relations*, Academic Press Inc., New York.
- Fodor, Jerry A. and Janet Dean Fodor. 1980. Functional structure, quantifiers, and meaning postulates. *Linguistic Inquiry* 11(4): 759–770.
- Foley, William A and Jr Van Valin, Robert D. 1984. *Functional syntax and universal grammar*. Cambridge University Press, Cambridge.
- Forker, Diana. 2014. A Canonical Approach to the Argument/Adjunct Distinction. *Linguistic Discovery* 12(2).
- Gaeta, Livio and Davide Ricca. 2006. Productivity in Italian word formation: A variable-corpus approach. *Linguistics* 44(1): 57–89.
- Goldberg, Adele E. 1995. *Constructions. A Construction Grammar Approach to Argument Structure*. University Of Chicago Press, Chicago/London.
- Goldberg, Adele E. and Farrell Ackerman. 2001. The pragmatics of obligatory adjuncts. *Language* 77(4): 798–814.
- Goodall, Grant. 2021. *The Cambridge handbook of experimental syntax*. Cambridge University Press, Cambridge.
- Gries, Stefan Th. and Dagmar Divjak. 2009. Behavioral profiles: A corpus-based approach to cognitive semantic analysis. In Vyvyan Evans and Stéphanie Pourcel, editors, *New Directions in Cognitive Linguistics*, John Benjamins Publishing Company, Amsterdam/Philadelphia, 57–75.
- Grimshaw, Jane. 1990. *Argument Structure*. MIT Press, Cambridge, MA.
- Grimshaw, Jane and Sten Vikner. 1993. Obligatory adjuncts and the structure of events. In Eric Reuland and Werner Abraham, editors, *Knowledge and language vol 2*, Springer, Dordrecht, 143–157.

- Grodner, Daniel, Edward Gibson, and Duane Watson. 2005. The influence of contextual contrast on syntactic processing: Evidence for strong-interaction in sentence comprehension. *Cognition* 95(3): 275–296.
- Haegeman, Liliane. 1994. *Introduction to government and binding theory*. Blackwell, Oxford.
- Haiden, Martin. 2005. *Theta theory*. Mouton de Gruyter, Berlin and New York.
- Harris, Alice and Lyle Campbell. 1995. *Historical Syntax in Cross-Linguistic Perspective*. Cambridge University Press, Cambridge.
- Harris, Randy Allen. 1993. *The Linguistics Wars*. Oxford University Press, Oxford.
- Haspelmath, Martin. 1989. From purposive to infinitive — a universal path of grammaticization. *Folia Linguistica Historica* 23(Historica vol. 10,1-2): 287–310.
- Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In M. Tomasello, editor, *The New Psychology of Language*, Lawrence Erlbaum Associates., Mahwah, 211–242.
- Haspelmath, Martin. 2008. Creating economical morphosyntactic patterns. In *Linguistic universals and language change*, Oxford University Press, Oxford, 185–214.
- Heine, Bernd. 1990. The dative in Ik and Kanuri. In William A. Croft, Suzanne Kemmer, and Keith Denning, editors, *Studies in Typology and Diachrony*, John Benjamins Publishing Company, Amsterdam and Philadelphia, 129–149.
- Heine, Bernd. 2008. Evolution of case systems. In Andrej Malchukov and Andrew Spencer, editors, *The Oxford Handbook of Case*, Oxford University Press, Oxford, 458–469.
- Heine, Bernd, Ulrike Claudi, and Friederike Hünemeyer. 1991. *Grammaticalization: A Conceptual Framework*. Chicago University Press, Chicago.
- Heine, Bernd and Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge University Press, Cambridge.
- Hietam, Katrin. 2003. *Definiteness and Grammatical Relations in Estonian*. Ph.d, University of Manchester.
- Hockett, C. F. 1955. *A manual of phonology*. Waverly Press, Baltimore.
- Hoekstra, Teun. 2004. *Arguments and Structure: Studies on the Architecture of the Sentence*. Mouton de Gruyter, Berlin/Boston.
- Hole, Daniel. 2015. Arguments and adjuncts. In Tibor Kiss and Artemis Alexiadou, editors, *Syntax - Theory and Analysis. An International Handbook, vol 2*, De Gruyter Mouton, Berlin/Munich/Boston.
- Hopper, Paul J. and Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2): 251–299.
- Hopper, Paul J. and Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge University Press, Cambridge.
- Huumo, Tuomas and Liina Lindström. 2014. Partitives across constructions: on the range of uses of the Finnish and Estonian “partitive subjects”. In Silvia Luraghi and Tuomas Huumo, editors, *Partitive cases and related categories*, De Gruyter Mouton, Berlin, 153–176.
- Iggesen, Oliver A. 2013. Number of Cases. In Matthew S. Dryer and Martin Haspelmath, editors, *The World Atlas of Language Structures Online*, Max Planck Institute for Evolutionary Anthropology, Leipzig.
- Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar*. MIT Press, Cambridge, MA.
- Jackendoff, Ray. 1990. *Semantic structures*. MIT Press, Cambridge, MA.
- Jackendoff, Ray S. 1976. Toward an explanatory semantic representation. *Linguistic Inquiry* 7: 89–150.

- Jacobs, Joachim. 1994. *Kontra valenz*. Wissenschaftlicher Verlag Trier, Trier.
- Jakobson, Roman. 1936. *Beitrag zur allgemeinen Kasuslehre: Gesamtbedeutung der russischen Kasus. Selected Writings II.* Mouton, The Hague.
- Janda, Laura A. and Olga Lyashevskaya. 2011. Grammatical profiles and the interaction of the lexicon with aspect, tense, and mood in Russian. *Cognitive Linguistics* 22(4): 719–763.
- Jespersen, Otto. 1924. *The philosophy of grammar*. George Allen and Unwin Ltd, London.
- Jónsson, Jóhannes Gíslí. 2003. Not so quirky: On subject case in Icelandic. *New perspectives on case theory* 127–163.
- Joos, Martin. 1950. Description of language design. *Journal of the Acoustical Society of America* 22: 701–708.
- Kaalep, Heiki-Jaan. 2018. Parallel forms in Estonian finite state morphology. *Proceedings of the 4th International Workshop for Computational Linguistics for Uralic Languages* 139–153.
- Karlsson, Fred. 1986. Frequency considerations in morphology. *STUF - Language Typology and Universals* 39(1-4): 19–28.
- Kasik, Reet. 2015. *Sõnamoodustus (Word Formation)*. Tartu University Press, Tartu.
- Kask, Arnold. 1960. Eesti keele lauseliigendusest ja lauseliikmeist. *Keel ja Kirjandus* 4: 237–243.
- Kennison, Shelia M. 1999. Processing Agentive By -Phrases in Complex Event and Nonevent Nominals. *Linguistic Inquiry* 30(3): 502–508.
- Kennison, Shelia M. 2002. Comprehending noun phrase arguments and adjuncts. *Journal of Psycholinguistic Research* 31(1): 65–81.
- Kettunen, Lauri. 1924. *Lauseliikmed eesti keeles (Parts of Speech in Estonian)*. Eesti Kirjanduse Seltsi kirjastus, Tartu.
- Kim, Jaegwon. 1969. Events and their descriptions: Some considerations. In Nicholas Rescher, editor, *Essays in honor of Carl G. Hempel*, Springer, New York, 198–215.
- Kittilä, Seppo. 2008. Case and the typology of transitivity. In Andrej Malchukov and Andrew Spencer, editors, *The Oxford Handbook of Case*, Oxford University Press, Oxford, 356–369.
- Kittilä, Seppo, Katja Västi, and Jussi Ylikoski. 2011. Introduction to Case, Animacy and Semantic Roles. In *Case, Animacy and Semantic Roles*, John Benjamins Publishing Company, Amsterdam.
- Klaas, Birute. 1988. Indirektne objekt (Indirect Object). *Keel ja Kirjandus* 1: 1–8.
- Klaas-Lang, Birute. 1999. No Dependence of the object case on the semantics of the verb in Estonian, Finnish and Lithuanian. In *Estonian: Typological studies III*, Tartu University Press, Tartu, 47–83.
- Klavan, Jane. 2012. *Evidence in linguistics: corpus-linguistic and experimental methods for studying grammatical synonymy*. Ph.d, University of Tartu.
- Klavan, Jane and Dagmar Divjak. 2016. The cognitive plausibility of statistical classification models: Comparing textual and behavioral evidence. *Folia Linguistica* 50(2): 355–384.
- Klavan, Jane, Kaisa Kesküla, and Laura Ojava. 2011. The division of labour between synonymous locative cases and adpositions: The Estonian adessive and the adposition peal ‘on’. In *Case, animacy and semantic roles*, John Benjamins Publishing Company, Amsterdam/Philadelphia, 113–134.

- Klotz, Michael. 2000. *Grammatik und lexik. Studien zur syntagmatik englischer verben*. Stauffenburg, Tübingen.
- Knoeferle, Pia, Boukje Habets, Matthew W. Crocker, and Thomas F. Münte. 2008. Visual scenes trigger immediate syntactic reanalysis: Evidence from ERPs during situated spoken comprehension. *Cerebral Cortex* 18(4): 789–795.
- Koenig, Jean Pierre, Gail Mauner, and Breton Bienvenue. 2003. Arguments for adjuncts. *Cognition* 89(2): 67–103.
- Konieczny, Lars, Barbara Hemforth, Christoph Scheepers, and Gerhard Strube. 1997. The Role of Lexical Heads in Parsing: Evidence from German. *Language and Cognitive Processes* 12(2-3): 307–348.
- Kont, Karl. 1963. *Käändsõnaline objekt läänemeresoome keeltes (Nominal object in Finnic languages)*. Eesti NSV Teaduste Akadeemia, Tallinn.
- Kratzer, Angelika. 2004. Telicity and the meaning of objective case. In Jacqueline Guéron and Jacqueline Lecarme, editors, *The Syntax of Time*, MIT Press, Cambridge, MA.
- Kulikov, Leonid. 2008. Evolution of case systems. In Andrej Malchukov and Andrew Spencer, editors, *The Oxford Handbook of Case*, Oxford University Press, Oxford, 439–457.
- Kurby, Christopher A. and Jeffrey M. Zacks. 2008. Segmentation in the perception and memory of events. *Trends in Cognitive Sciences* 12(2): 72–79.
- Kure, Kristjan. 1959. Eesti keele lauseliigenduse alustest. *Keel ja Kirjandus* 1: 40–49.
- Kuryłowicz, Jerzy. 1964. *The Inflectional Categories of Indo-European*. Carl Winter Universitätsverlag, Heidelberg.
- Kılıçaslan, Yılmaz and Gurkan Tuna. 2015. Linking FrameNet and Natural Languages via Thematic Role Structures. *International Journal of Languages, Literature and Linguistics* 1(3): 158–163.
- Labov, William. 1973. The boundaries of words and their meanings. In C.-J. Bailey and R. W. Shuy, editors, *New ways of analyzing variation in English*, Georgetown University Press, Washington DC, 340–373.
- Lakoff, George. 1966. Stative adjectives and verbs in English. In Anthony G. Oettinger, editor, *Mathematical linguistics and automatic translation*, Aiken Computation Laboratory, Harvard University, Cambridge, MA.
- Lakoff, George. 1970. Linguistics and natural logic. *Synthese* 22(1-2): 151–271.
- Lakoff, George. 1973. Fuzzy grammar and the performance/competence terminology game. *UC Berkeley Previously Published Works*.
- Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. University Of Chicago Press, Chicago and London.
- Lakoff, George. 1999. Cognitive models and prototype theory. In Eric Margolis and Stephen Laurence, editors, *Concepts*, MIT Press, Cambridge, MA, 391–421.
- Larjavaara, Matti. 1991. Aspektuaalisen objektin synty. *Virtittäjä* 372–408.
- Lees, Aet. 2004. The Case of the object in early Estonian and Finnish texts. *Proceedings of the 2004 Conference of the Australian Linguistics Society*.
- Lehari, Valdar. 1959. Viimaseid süntaksialaseid kirjutisi lugedes. *Keel ja Kirjandus* 6: 356–362.
- Lehmann, Christian. 1985. Synchronic Variation and Diachronic Change. *Lingua e Stile* 20: 303–318.
- Lestrade, Sander. 2010. *The space of case*. Ph.d dissertation, Radboud University.
- Levin, Beth and Malka Rappaport-Hovav. 2005. *Argument Realization*. Cambridge University Press, Cambridge.

- Lindström, L., M. Kalmus, A. Klaus, L. Bakhoff, and K. Pajusalu. 2009. Ainsuse 1. isikule viitamine eesti murretes. *Emakeele Seltsi aastaraamat* 54: 159–185.
- Lindström, Liina. 2005. *Finiitverbi asend lauses. Sõnajärg ja seda mõjutavad tegurid suulises eesti keeles*. Ph.d, University of Tartu.
- Lindström, Liina and Kristel Uihoaed. 2017. Syntactic variation in 'need'-constructions in Estonian dialects. *Nordic Journal of Linguistics* 40(3): 313–349.
- Lindström, Liina and Virve Anneli Vihman. 2017. Who needs it? Variation in experiencer marking in Estonian 'need'- constructions. *Journal of Linguistics* 53(4): 789–822.
- MacDonald, Maryellen C., Neal J. Pearlmutter, and Mark S. Seidenberg. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101(4): 676–703.
- Maienborn, Claudia. 2003. Event-internal modifiers: Semantic underspecification and conceptual interpretation. In Ewald Lang, Claudia Maienborn, and Cathrine Fabricius-Hansen, editors, *Modifying adjuncts*, Mouton De Gruyter, Berlin, 475–509.
- Maienborn, Claudia. 2008. On Davidsonian and Kimian states. In Ileana Comorovski and Klaus von Heusinger, editors, *Existence: Semantics and syntax*, Springer, Dordrecht, volume 84, 107–130.
- Maienborn, Claudia. 2019. Events and States. In Robert Truswell, editor, *The Oxford Handbook of Event Structure*, Oxford University Press, Oxford, January, 49–89.
- Malchukov, Andrej and Bernard Comrie. 2015. *Valency Classes in the World's Languages, vol 1*. Walter de Gruyter, Berlin/Boston.
- Malchukov, Andrej and Peter De Swart. 2008. Differential case marking and actancy variations. In Andrej Malchukov and Andrew Spencer, editors, *The Oxford Handbook of Case*, Oxford University Press, Oxford.
- Malchukov, Andrej L. 2005. Case pattern splits, verb types and construction competition. In *Competition and Variation in Natural Languages*, Elsevier Science, 73–117.
- Marantz, Alec. 2013. Verbal argument structure: Events and participants. *Lingua* 130: 152–168.
- Matsumura, Kazuto. 1994. Is the Estonian adessive really a local case? *Journal of Asian and African Studies* 46-47: 223–235.
- Mathews, P. H. 1981. *Syntax*. Cambridge University Press, Cambridge.
- Mathews, Peter. 2008. The scope of valency in grammar. In Thomas Herbst and Katrin Götz-Votteler, editors, *Valency: Theoretical, Descriptive and Cognitive Issues*, Mouton de Gruyter, Berlin and New York, 3–14.
- Maurer, Gail and Jean Pierre Koenig. 1999. Lexical encoding of event participant information. *Brain and Language* 68(1-2): 178–184.
- Maurer, Gail and Jean Pierre Koenig. 2000. Linguistic vs. Conceptual Sources of Implicit Agents in Sentence Comprehension. *Journal of Memory and Language* 43(1): 110–134.
- Maurer, Gail, Alissa Melinger, Jean-Pierre Koenig, and Breton Bienvenue. 2002. When is schematic participant information encoded? Evidence from eye-monitoring. *Journal of Memory and Language* 47: 386–406.
- Metslang, Ann. 2014. Ainsuse pika ja lühikese sisseütleva valiku olenemine morfofonoloogilistest tunnustest - korpusanalüüs. *Emakeele Seltsi Aastaraamat* 60(1): 127–147.
- Metslang, H, K Habicht, and K Pajusalu. 2017. Where Do Polar Question Markers Come From?
- Metslang, Helena. 2008. Sihitisesarnased määrused: kui sarnased? In Helle Metslang and Riho Grünthal, editors, *Lähivertailuja 18*, Suomalais-ugrilainen seura, Helsinki, 108–143.

- Metslang, Helena. 2012. On the case-marking of existential subjects in Estonian. *SKY Journal of Linguistics* 25(2012): 151–204.
- Mihkla, Karl. 1959. Mõningaid märkmeid lauseliikmete ja lauseanalüüsi kohta. *Keel ja Kirjandus* 3: 171–176.
- Mihkla, Karl, Lehte Rannut, Elli Riikoja, and Aino Admann. 1974. *Eesti keele lauseõpetuse põhijooned I (Foundations of Estonian Syntax I)*. Valgus, Tallinn.
- Miljan, Merilin. 2009. *Grammatical case in Estonian*. Ph.d, University of Edinburgh.
- Müller, Stefan. 2020. *Grammatical theory. From transformational grammar to constraint-based approaches*. Language Science Press, Berlin, 2nd edition.
- Murphy, G. L. 2002. *The big book of concepts*. MIT Press, Cambridge, MA.
- Nemvalts, Peep. 2000. *Aluse sisu ja vorm*. Eesti Keele Sihtasutus, Tallinn.
- Neustupny, J. V. 1966. On the analysis of linguistic vagueness. *Travaux linguistiques de Prague* 2: 39–51.
- Newmeyer, Frederick, J. 2010. On comparative concepts and descriptive categories: A reply to Haspelmath. *Language* 86(3): 688–695.
- Nichols, Johanna. 1983. On Direct and Oblique Cases. In *Proceedings of the Ninth Annual Meeting of the Berkeley Linguistics Society (1983)*. Berkeley, 170–192.
- Nichols, Johanna. 1984. Direct and oblique objects in Chechen-Ingush and Russian. In Frans Plank, editor, *Objects. Towards a theory of grammatical relations*, Academic Press Inc., London, 183–209.
- Nichols, Johanna and Alan Timberlake. 1991. Grammaticalization as retextualization. In Elizabeth Closs Traugott and Bernd Heine, editors, *Approaches to grammaticalization*, John Benjamins Publishing Company, Amsterdam and Philadelphia, 129–146.
- Nurka, Maiu. 2014. *Eesti keele elatiivi semantika kognitiivse keeleteaduse vaatenurgast (Semantics of Estonian Elative from a Cognitive Linguistics Perspective)*. Ma thesis, University of Tartu.
- Ogren, David. 2015. Differential object marking in Estonian: Prototypes, variation, and construction-specificity1. *SKY Journal of Linguistics* 28(2015): 277–312.
- Ogren, David. 2018. *Object case variation in Estonian da-infinitive constructions*. Ph.d, University of Tartu.
- Onishi, Masayuki. 2001. Non-canonically marked subjects and objects: Parameters and properties. In Alexandra Y. Aikhenvald, R.M.W Dixon, and Masayuki Onishi, editors, *Non-canonical marking of subjects and objects*, John Benjamins, Amsterdam and Philadelphia, 1–53.
- Perek, Florent. 2015. *Argument Structure in Usage-Based Construction Grammar*. John Benjamins Publishing Company, Amsterdam.
- Pesetsky. 1982. *Paths and categories*. Ph.d thesis, MIT.
- Pilvik, Maarja-Liisa. 2019. Assessing the productivity of the Estonian deverbal suffix -mine in five registers of Estonian. *SKY Journal of Linguistics* 32: 75–103.
- Pinker, Steven. 1999. *Words and Rules: The Ingredients of Language*. Basic Books.
- Primus, Beatrice. 1999. *Cases and thematic roles*. De Gruyter, Berlin.
- Proos, Mariann. 2021. *Meaning and usage of Estonian experience perception verbs*. Ph.d dissertation, University of Tartu.
- Radford, A. 1988. *Transformational Grammar: A First Course*. Cambridge University Press, Cambridge.
- Rajandi, Henno and Helle Metslang. 1979. *Määramata ja määratud objekt*. Valgus, Tallinn.
- Rákosi, György. 2006. *Dative experiencer predicates in Hungarian*. LOT, Utrecht.

- Rätsep, Huno. 1959. Põhimõttelist seost süntaksialase diskussiooniga. *Keel ja Kirjandus* 12: 726–730.
- Rätsep, Huno. 1978. *Eesti keele lihtlausete tüübid*. Valgus, Tallinn.
- Rätsep, Huno. 1979. *Eesti keele ajalooline morfoloogia II (Historical Morphology of Estonian II)*. Tartu National University, Tartu.
- Rice, Sally and Kaori Kabata. 2007. Crosslinguistic grammaticalization patterns of the allative. *Linguistic Typology* 11(3): 451–514.
- Rissman, Lilia and Asifa Majid. 2019. Thematic roles: Core knowledge or linguistic construct? *Psychonomic Bulletin and Review* 26(6): 1850–1869.
- Rissman, Lilia, Kyle Rawlins, and Barbara Landau. 2015. Using instruments to understand argument structure: Evidence for gradient representation. *Cognition* 142: 266–290.
- Robins, R. H. 1990. *A short history of linguistics*. Longman, London, 3rd edition.
- Rosch, Eleanor. 1978. Principles of categorization. In E. Rosch and B. B. Lloyd, editors, *Cognition and categorization*, Erlbaum, Hillsdale NJ, 27–48.
- Rosch, Eleanor and Carolyn B. Mervis. 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7(4): 573–605.
- Rosch, Eleanor H. 1973. On the internal structure of perceptual and semantic categories. In Timothy E. Moore, editor, *Cognitive development and the acquisition of language*, Academic Press, New York, 111–144.
- Rosch, Eleanor H., Carol Simpson, and R. Scott Miller. 1976. Structural bias of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance* 2(4): 491–502.
- Ross, John Robert. 1972. The category squish: Endstation Hauptwort. In Paul M. Peranteau, Judith N. Levi, and Gloria C. Phares, editors, *Papers from the Eighth Regional Meeting of the Chicago Linguistic Society*. Chicago Linguistic Society, Chicago, volume 8, 316–328.
- Ross, Kristiina. 1997. Kohakääänded Georg Mülleri ja Heinrich Stahli eesti keeles. In *Pühendusteos Huno Rätsepale. Tartu Ülikooli eesti keele õppetooli toimetised 7*, Tartu Ülikooli Kirjastus, Tartu, 184–201.
- Rothmayr, Antonia. 2009. *The Structure of Stative Verbs*. John Benjamins, Amsterdam and Philadelphia.
- Saks, K. 1959. Kaudsihitise kaitseks. *Keel ja Kirjandus* 8(485–487).
- Saluveer, Viktor. 1958. *Sise- ja väliskohakäänete tähenduslikud funktsioonid eesti keeles*. Ma thesis, Tartu National University.
- Schlesinger, I.M. 1995. *Cognitive Space and Linguistic Case*. Cambridge University Press, Cambridge.
- Schütze, Carson T. 1995. Attachment and Argumenthood. *MIT Working Papers in Linguistics* 26 (Dir 9113607): 95–151.
- Schütze, Carson T. and Edward Gibson. 1999. Argumenthood and English Prepositional Phrase Attachment. *Journal of Memory and Language* 40(3): 409–431.
- Siiman, Ann. 2015. Ainsuse sisseütleva vormi valiku seos morfosüntaktiliste ja semantiliste tunnustega: Materjali ning meetodi sobivus korpusanalüüsiks. *Emakeele Seltsi Aastaraamat* 61(1): 207–232.
- Siiman, Ann. 2019. *Vormikasutuse varieerumine ja selle põhjused osastava ja sisseütleva käände näitel*. Ph.d, University of Tartu.
- Silverstein, Michael. 1976. Hierarchy of features and ergativity. In R.M.W. Dixon, editor, *Grammatical categories in Australian languages*, Australian Institute of Aboriginal Studies, Canberra, 112–171.

- Sirola-Belliard, Maija. 2011. Is there a future for the Finnish comitative? Arguments against the putative synonymy of the comitative case -ine and the postposition kanssa. In Seppo Kittilä, Jussi Ylikoski, and Katja Västi, editors, *Case, animacy and semantic roles*, John Benjamins Publishing Company, Amsterdam and Philadelphia.
- Somers, Harold L. 1984. On the validity of the complement-adjunct distinction in valency grammar. *Linguistics* 22(4): 507–530.
- Speer, Shari R. and Charles Clifton. 1998. Plausibility and argument structure in sentence comprehension. *Memory and Cognition* 26(5): 965–978.
- Spivey-Knowlton, Michael and Julie C. Sedivy. 1995. Resolving attachment ambiguities with multiple constraints. *Cognition* 55(3): 227–267.
- Stamenković, Tusan and Miloš Tasić. 2013. English Verbs of Motion and Prototype Theory. *British and American Studies* (October).
- Svenonius, Peter. 2005. Case alternations in the Icelandic passive. *Passives and impersonals in European languages* 1–23.
- Svorou, Soteria. 1994. *The grammar of space*. John Benjamins Publishing Company, Amsterdam and Philadelphia.
- Tamm, Anne. 2004. *Relations between Estonian verbs, aspect, and case*. Ph.d thesis, Eötvös Loránd University.
- Taremaa, Piia. 2017. *Attention meets language: a corpus study on the expression of motion in Estonian*. Ph.d, University of Tartu.
- Taylor, John R. 1995. *Linguistic categorization: Prototypes in linguistic theory*. Clarendon Press, Oxford.
- Tenny, Carol. 1992. The aspectual interface hypothesis. In Ivan. A. Sag and Anna Szabolcsi, editors, *Lexical matters*, CSLI Publications, Stanford, 1–27.
- Tesnière, Lucien. 2015. *Elements of Structural Syntax*. John Benjamins Publishing Company, Amsterdam/Philadelphia.
- Torn, Reeli. 2006. Oblique dependents in Estonian: an LFG perspective. In Miriam Butt and Tracy Holloway King, editors, *Eleventh International Lexical Functional Grammar Conference 2006*. CSLI Publications, Stanford, 504–515.
- Traxler, Matthew J. and Kristen M. Tooley. 2007. Lexical mediation and context effects in sentence processing. *Brain Research* 1146(1): 59–74.
- Tsunoda, Tasaku. 1981. Split case-marking patterns in verb-types and tense/aspect/mood. *Linguistics* 19: 389–438.
- Tutunjian, Damon and Julie E. Boland. 2008. Do we need a distinction between arguments and adjuncts? Evidence from psycholinguistic studies of comprehension. *Linguistics and Language Compass* 2(4): 631–646.
- Vääri, Eduard. 1959. Keele uurimise meetodeist, ”kaudsihitistest” ja öeldistäitest. *Keel ja Kirjandus* 4: 228–232.
- Vainik, Ene. 1995. *Eesti keele väliskohakäänete semantika kognitiivse grammatika vaatenurgast (Semantics of Estonian external cases from the perspective of Cognitive Grammar)*. Eesti TA Eesti Keele Instituut, Tallinn.
- Vaiss, Natalia. 2021. Verbi transitiivsuse kontiinumist eesti keeles (Verbal transitivity continuum in Estonian). *Emakeele Seltsi Aastaraamat* 66: 344–386.
- Van Valin, Robert and Randy LaPolla. 1997. *Syntax: Structures, Meaning, and Function*. Cambridge University Press, Cambridge.
- Van Valin, Robert D. 2005. *Exploring the syntax-semantics interface*. Cambridge University Press, Cambridge.

- Van Valin, Robert D., Jr. 1990. Semantic parameters of split intransitivity. *Language* 66(2): 221–260.
- Västi, Katja. 2011. A case in search of an independent life: The semantics of the initial allative in a Finnish verbless construction. In Seppo Kittilä, Jussi Ylikoski, and Katja Västi, editors, *Case, animacy and semantic roles*, John Benjamins Publishing Company, Amsterdam and Philadelphia, 65–112.
- Veismann, Ann, Mati Ereht, and Helle Metslang. 2017. Määrus (Adverbial). In Mati Ereht and Helle Metslang, editors, *Eesti keele süntaks*, Tartu University Press, Tartu, 300–375.
- Vendler, Zeno. 1957. Verbs and times. *The Philosophical Review* 66(2): 143–160.
- Verkuyl, Henk J. 1999. *Aspectual issues. Studies on time and quantity*. University of Chicago Press, Chicago.
- Vihman, Virve Anneli. 2015. Pick it up: A look at referential devices in Estonian child-directed speech. *Eesti ja Soome-Ugri Keeleteaduse Ajakiri* 6(2): 63–83.
- Viht, Annika and Külli Habicht. 2019. *Eesti keele sõnamuutmine (Estonian Inflectional Morphology)*. Tartu University Press, Tartu.
- Viitso, Tiit-Rein. 2003. Structure of the Estonian language. Phonology, morphology and word formation. In Mati Ereht, editor, *Estonian Language*, Estonian Academy of Sciences, Tallinn, 9–92.
- Viks, Ülle. 1982. *Klassifikatoorne morfoloogia. Noomen..* Eesti NSV Teaduste Akadeemia, Tallinn.
- Viks, Ülle. 1992. *Väike vormisõnastik I: Sissejuhatus ja grammatika*. Eesti Teaduste Akadeemia, Tallinn.
- Wiedemann, Ferdinand Johann. 1875. *Eesti keele grammatika*. Eesti Teaduste Akadeemia, Tallinn.
- Williams, Alexander. 2015. *Arguments in syntax and semantics*. Cambridge University Press, Cambridge.
- Wunderlich, Dieter and Renate Lakämper. 2001. On the interaction of structural and semantic case. *Lingua* 111(4-7): 377–418.
- Yamashita, Hiroko. 1995. Verb argument information used in a pro-drop language: an experimental study in Japanese. *Journal of Psycholinguistic Research* 24: 333–347.
- Ylikoski, Jussi. 2011. A survey of the origins of directional case suffixes in European Uralic. In *Case, animacy and semantic roles*, John Benjamins Publishing Company, Amsterdam/Philadelphia, 235–282.
- Zadeh, L. A. 1965. Fuzzy sets. *Information and Control* 8: 338–353.
- Zaenen, Annie and Joan Maling. 1984. Unaccusative, passive, and quirky case. In Joan Maling and Annie Zaenen, editors, *Modern Icelandic Syntax*, Brill, Leiden, Syntax and Semantics 24, 137–152.
- Zeldes, Amir. 2012. *Productivity in argument selection: From morphology to syntax*. De Gruyter Mouton, Berlin and Boston.
- Zipf, G. K. 1949. *Human behavior and the principle of least effort. An introduction to human ecology*. Hafner Publishing Company, London and New York.

## CURRICULUM VITAE

**Name:** Mari Aigro  
**Date of birth:** August 9, 1990  
**Citizenship:** Estonian  
**E-mail:** mari.aigro@ut.ee  
**Address:** Jakobi 2-430, Institute of Estonian and General Linguistics,  
University of Tartu, 51005, Tartu, Estonia

### Education:

2017– University of Tartu, Institute of Estonian and General Linguistics, Ph.D studies  
2016–2017 University of Cambridge, MPhil Theoretical and Applied Linguistics (*with distinction*)  
2013–2016 University of Aberdeen, MA Language and Linguistics (*first degree*)  
1997–2009 Mustamäe Secondary School (*with distinction*)

### Employment history:

2019– University of Tartu, Institute of Estonian and General Linguistics, Junior Research Fellow

### Publications:

Aigro, M; Vihman, V. (*in press*) Oblique complements in Estonian: A corpus perspective. *Journal of Linguistics*.

Aigro, M; Vihman, V. (*in review*) Overabundance in Estonian noun paradigms.

Aigro, M; Vihman, V. (*submitted*) Jumalast lõbus: Evidence of elative and allative grammaticalisation from spoken language.

Proos, M; Aigro, M. (*submitted*) Concreteness ratings for 36,000 Estonian words.

Aigro, M; Proos, M (2022) Konkreetushinnangute kogumine eestikeelsetele sõnadele [Collecting Concreteness Ratings for Estonian Words]. *Eesti Rakenduslingvistika Aastaraamat* 18. 5–20

Aigro, M (2020) Polar Question Particles and Their Sources: A Semantic Approach to Grammaticalization. *Transactions of the Philological Society* 118 (2). 237–253.

Aigro, M (2018) Kvantitatiivne pööre – vana tüli uues kuues [Quantitative Shift: An Old Quarrel Renewed]. *Keel ja Kirjandus* 8–9. 747–764.

## ELULOOKIRJELDUS

**Nimi:** Mari Aigro  
**Sünniaeg:** 9. august, 1990  
**Kodakondsus:** eestlane  
**E-mail:** mari.aigro@ut.ee  
**Aadress:** Jakobi 2-430, Eesti ja üldkeeleteaduse instituut, Tartu  
Ülikool, 51005, Tartu

### Haridus:

2017– Tartu Ülikool, Eesti ja üldkeeleteaduse instituut, doktoriõpe (üldkeeleteadus)  
2016–2017 Cambridge'i Ülikool (UK), MPhil Theoretical and Applied Linguistics (magistrikraad kiitusega, *with distinction*)  
2013–2016 Aberdeeni Ülikool (UK), MA Language and Linguistics (bakalaureusekraad kiitusega, *first degree*)  
1997–2009 Mustamäe Gümnaasium, keskkharidus (inglise keele ja reaalinete süvaõpe, kuldmedal)

### Teenistuskäik:

2019– Tartu Ülikool, Eesti ja üldkeeleteaduse instituut, nooremteadur

### Publikatsioonid:

Aigro, M; Vihman, V. (*ilmumas*) Oblique complements in Estonian: A corpus perspective. *Journal of Linguistics*.

Aigro, M; Vihman, V. (*retsenseerimisel*) Overabundance in Estonian noun paradigms.

Aigro, M; Vihman, V. (*esitatud*) Jumalast lõbus: Evidence of elative and allative grammaticalisation from spoken language.

Proos, M; Aigro, M. (*esitatud*) Concreteness ratings for 36,000 Estonian words.

Aigro, M; Proos, M (2022) Konkreet-sushinnangute kogumine eestikeelsetele sõnadele. *Eesti Rakenduslingvistika Aastaraamat* 18. 5–20

Aigro, M (2020) Polar question particles and their sources: A semantic approach to grammaticalization. *Transactions of the Philological Society* 118 (2). 237–253.

Aigro, M (2018) Kvantitatiivne pööre – vana tüli uues kuues. *Keel ja Kirjandus* 8–9. 747–764.

## DISSERTATIONES LINGUISTICAE UNIVERSITATIS TARTUENSIS

1. **Anna Verschik.** Estonian Yiddish and its contacts with coterritorial 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 sissejuhatus: 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 paran-dused. 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 argi-vestluses. 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 Tälberga.** 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.
42. **Helen Hint.** From full phrase to zero: a multifactorial, form-specific and crosslinguistic analysis of Estonian referential system. Tartu, 2021, 271 p.
43. **Anton Malmi.** The production of Estonian palatalization by Estonian and Russian speakers. Tartu, 2022, 180 p.