Humaniora

DISSERTATIONES
LINGUISTICAE
UNIVERSITATIS
TARTUENSIS
15

JANE KLAVAN

Evidence in linguistics: corpus-linguistic and experimental methods for studying grammatical synonymy

TARTU ÜLIKOOL
1632
UNIVERSITAS TARTUENSIS

DISSERTATIONES LINGUISTICAE UNIVERSITATIS TARTUENSIS
15

DISSERTATIONES LINGUISTICAE UNIVERSITATIS TARTUENSIS
15

JANE KLAVAN

Evidence in linguistics: corpus-linguistic and experimental methods for studying grammatical synonymy

UNIVERSITY OF TARTU
PRESS
1632

University of Tartu, Institute of Estonian and General Linguistics

Dissertation accepted for the commencement of the degree of Doctor of Philosophy on October 18th, 2012 by the Committee of the Institute of Estonian and General Linguistics, Faculty of Philosophy, University of Tartu

Supervisor: dots. Ilona Tragel, University of Tartu
dr. Dagmar Divjak, University of Sheffield

Opponent: prof. Ewa Dąbrowska, Northumbria University

Commencement: November 29th, 2012 at 14.15, room 139 in University main building, Ülikooli 18, Tartu

This study has been supported by the European Social Fund

ISSN 1406–5657
ISBN 978–9949–32–165–0 (print)
ISBN 978–9949–32–166–7 (pdf)

University of Tartu Press
www.tyk.ee
Tellimus nr. 559

PREFACE

Like so many, if not all, great things in life, I became a linguist by accident. Back in high school I never thought I would be writing a doctoral dissertation about Estonian locative cases and adpositions. I have always had a special bond with English (some might call it an obsession), but if anything, I was absolutely sure that I would one day write a book about Elizabeth I, Queen of England. This obsession with English history and language persists to this day, but somehow I found my way to the Department of General Linguistics after doing my BA and MA in English Language and Linguistics.

I guess it all began when, as a second year BA student, I shyly approached our English Grammar lecturer, Reeli Torn-Leesik, and asked her if she would like to play tennis with me. She kindly agreed to beat me once a week on a tennis court and through our mutual love for tennis I also discovered that I shared Reeli's passion for linguistics. Thus, through fruit and vegetable metaphors (maybe not my proudest moment, albeit a vital one) and English spatial prepositions, I ended up writing a dissertation about grammatical synonymy. It was Reeli who, during my MA studies, encouraged me to contact Ilona Tragel and Ann Veismann once it became clear that I was treading the cognitive path in linguistics. I remember vividly the day I went to meet Ilona for the first time. We met at the University Café and I recall the slightly suspicious and intimidating look she gave me when I expressed my determination of becoming a cognitive linguist. Still, for all my sins, Ann and Ilona welcomed me with open arms and took me into the Tartu University cognitive linguistics in-crowd. Over the subsequent meetings, Ilona casually mentioned the topic of Estonian locative cases, and once it registered with me what a splendid research topic it was, I instantly fell in love with it. The fact that I have managed to put this story between covers is a sign that I was endowed with a topic that suited me down to the ground. My only hope is that I have done this wonderful area of research justice with this humble contribution.

There are very many people I want to thank for their support. And I am sure they all know it themselves as well. So those who are not explicitly mentioned, know that my gratitude to you is out there, in the universe.

First and foremost I thank my supervisor Ilona Tragel. She has always known when to push the right buttons and if it weren't for her academic as well as personal skills to motivate, this dissertation would never have seen light. I appreciate enormously that she did not force me in the direction of research more close to her own heart and that she let me do my own thing (I am a Taurus, after all) and grow as an independent researcher. It must be hard to watch your "children" commit the same mistakes you have maybe made yourself already years ago, but it is personal experience that provides the best supervision. I am also extremely grateful for Ilona's endless patience. It must have been strenous to put up with all the number crunching (the tables and figures) that riddles this dissertation. She has done so much for me that I will remain indebted to her forever.

I also thank Dagmar Divjak, my co-supervisor in Sheffield. I am deeply grateful that she agreed to become my supervisor at the eleventh hour. She has been a great source of inspiration and her meticulousness and ingenuity in doing research is simply amazing. I am extremely fortunate that I got to know her both academically and personally. I appreciate enormously her prompt and detailed comments on my drafts, even if she had to deliver them at Sheffield Children's Hospital. It is her dedication and commitment to doing research that I honour the most. I have learned a lot from both of my supervisors and I owe them both a lot. I can only hope that one day I will partly be able to repay this debt by becoming as good a supervisor as Ilona and Dagmar have been to me.

I am grateful to many colleagues at Tartu and outside for their kindness and inspiration. I am especially thankful to Ann Veismann and Anni Jürine with whom I have been able to share my ideas and passion for adpositional semantics and doing linguistic experiments. Ann has always been there for me whenever I needed advice or a second opinion about something. And the many colourful adventures me and Anni have shared on international conferences merit a book-length description of their own. I also thank Reeli Tron-Leesik for showing me the way to linguistics and Kaja Kährik for lending me her collection on cognitive linguistics. My gratitude goes also to the colleagues and students who have throughout the years participated in the Cognitive Linguistics seminars at Tartu University. From among the many colleagues outside the alma mater, I want to single out Dylan Glynn. I will be forever grateful that he was willing to discuss my work whenever we met during a conference or a workshop and, of course, for introducing me to the wonderful world of R. His patience and help was greatly valued. My heart-felt thanks also go to my colleagues at the Department of English: Age Allas, Berk Vaher, Ene-Reet Soovik, Enn Veldi, Eva Rein, Ilmar Anvelt, Kärt Vahtramäe, Katiliina Gielen, Katri Sirkel, Krista Kallis, Krista Vogelberg, Leili Kostabi, Mariann Enno, Natalja Zagura, Pille Põiklik, Pilvi Rajamäe, Piret Kärtner, Piret Rääbus, Raili Marling, Reet Sool, and Ülle Türk.

I am also grateful to the reviewers of the dissertation – Ewa Dąbrowska and Maarten Lemmens – for sharing their valuable expertise in cognitive linguistics and providing useful comments. Naturally, the responsibility for any errors or omissions that remain is entirely mine.

As for official support, I have benefited from the Graduate School of Linguistics and Language Technology, the Graduate School of Linguistics, Philosophy and Semiotics, the Target Financed Research Programme "Language and meaning: semantics and grammar in a cognitive perspective" (SF0180056s08), and the DoRa and Kristjan Jaak programmes of SA Archimedes.

I would also like to express my thanks to everyone who tried to keep me from drowning into my thesis and for providing a life outside the university, at which they occasionally succeeded too well. Thank you to my good friends from the indoor hockey team, especially Margit Keerutaja and Reti Väärt. As Carl Robert Jakobson said "Terves kehas, terve vaim" (*In a healthy body, there is a healthy mind*). Thanks also to my very dear friend Elli Põder, who was

always willing to have a glass of wine or two whenever the world needed straightening out.

I am deeply grateful to my family. I am very blessed and lucky to have them in my life. Thank you to my sister Diana, my brother Ragnar, his wife Lily and their baby son Romer. They have always been there for me throughout my studies and have never complained when for yet another birthday or Christmas I gave them a book as a present. Of course, I am much indebted (quite literally) to my parents, Tiina and Dzintar, who have supported me throughout the years in too many ways. It is only thanks to their continuous encouragement to follow my heart that I have been able to enjoy the life in academia. I am forever grateful to them for raising me the way they did.

And finally, Paul, thank you for being part of my life. You are the other, practical and sane, half of me. I have sacrificed too many things too many times in order to achieve my academic and personal goals. It is now time that I make good by giving you back the love and affection you have given me.

CONTENTS

TABLES

FIGURES

ABBREVIATIONS

List of abbreviations used in the glosses

ADE	adessive
ALL	allative
COM	comitative
ESS	essive
GEN	genitive
ILL	illative
INE	inessive
INF	infinitive
NOM	nominative
PASS	passive
PL	plural
PRS	present
PRT	partitive
PST	past
PTCP	participle
SG	singular
SUP	supine

List of other abbreviations used in the dissertation

ade	adessive construction
BCE	Balanced Corpus of Estonian
CX/cx	construction
LM/lm	Landmark
LM_COMP	morphological complexity of the Landmark
LM_LENGTHSYLLOG	length of the Landmark phrase in syllables (logarithmically transformed)
LM_MOBILITY	mobility of the Landmark
lm_tr	Landmark precedes the Trajector
lm1	Landmark precedes the Trajector
lm2	Landmark follows the Trajector
MDCE	Morphologically Disambiguated Corpus of Estonian
TR/tr	Trajector
tr_lm	Trajector precedes the Landmark
TR_WC	word class of Trajector
WO/wo	word order
WO_LM	relative position between Trajector and Landmark

INTRODUCTION

Like so many other students of cognitive linguistics before me, I found my way to Langacker and the other founding fathers of cognitive linguistics through the study of spatial language. It started with the contrastive study of Estonian and English medial region adpositions for my MA thesis (Klavan 2008) and continued with the study of locative cases and adpositions – the topic of the present study. Although there are many things besides locatives that are important in cognitive linguistics, I agree with Langacker in that locatives are rather subtle and complex and "we don't understand them all that well yet, so there is much to do still" (Langacker 2007: 252). One of the aims of the present work is therefore to contribute to a fuller understanding of locatives and spatial language in general. In this study I look at two Estonian locative constructions that both express a support-relation – a spatial situation where one object is located on top of or is attached to another object. The present study focuses on the alternation between the synthetic adessive case construction and the analytic adpositional construction with *peal* 'on'. The focal point is on employing a selection of different methodological tools (both corpus and experimental methods are made use of) which are applicable for the study of alternations between synthetic and analytic forms and grammatical synonymy in general.

Examples 1 and 2 illustrate how Estonian has at least two different ways to express a situation where an object is located on top of another object (the support-relation). The first option for describing the location of a book on a table is to use the adessive case, as *laual* in example 1 (referred to as the synthetic locative construction). The second option is to use the adposition *peal* 'on', as *laua peal* in example 2 (the analytic locative construction).

(1) *Raamat* *on* ***laual***.
book.SG.NOM[1] be-PRS.3SG table.SG.ADE
'The book is **on the table**.'

(2) *Raamat* *on* ***laua*** ***peal***.
book.SG.NOM be-PRS.3SG table.SG.GEN on
'The book is **on the table**.'

The study of (grammatical) synonymy makes it possible to address an intriguing phenomenon in linguistics. Languages are said to be economical, and yet there are numerous instances of synonymy in every language. There seems to be a contradiction – on the one hand, language exhibits economy par excellence via polysemy, that is making a single linguistic unit express multiple meanings, and on the other hand we have synonymous linguistic units decreasing the expressive power of languages by allowing several linguistic units to convey more or less the same meaning. Even if two linguistic units do express one and the same function, they do it in different ways: they allow for a different

[1] The list of abbreviations (including those used in the glosses) is provided on page 17.

construal of the same situation. It is assumed that the two locative constructions, the synthetic adessive case and the analytic adposition *peal* 'on', are not in free variation. Given that the topic of alternation between synthetic locative cases and analytic adpositional constructions has not previously received systematic treatment in (Estonian) linguistics, the aim of the thesis is to present a list of variables that influence the language speaker's choice between the two constructions.

One of the key theoretical starting points of the present thesis is the cognitive linguistic assumption that grammar is meaningful. Elements of grammar, for example, locative cases and adpositions, have meanings in their own right. But meaning, of course, presents a cognitive-functionally oriented linguist with a conundrum. On the one hand, meaning is central to any linguistic analysis, but on the other hand it is an elusive phenomenon. Researchers are faced with the hurdles of what and how meaning can be researched. In this study, I look at what the language itself provides. The sentence structure surrounding a grammatical construction, and in particular the sets of nouns the locative case or the adposition occurs with, provide clues to that construction's meaning. Specifically, I look at how different variables, both semantic and morphosyntactic, interact and contribute to the subtle meaning differences exhibited by the two constructions. In studying the division of labour between the synthetic locative case and the analytic adpositional construction, I present a solution to one of the problems of synonymy research – is there an objective way to decide which variables distinguish best between semantically similar locative constructions.

The general theoretical framework of this study combines several approaches. It is usage-based, it proceeds from the premises of Cognitive Grammar and Construction Grammar, and it employs methodological pluralism (i.e. combining introspection, corpus and experimental data). The dissertation advocates linguistic empiricism, a label used by Arppe and Järvikivi (2007: 135). In such an approach, the everyday language used by speakers assumes the leading role. Theory follows from what the data present and not vice versa. The bulk of the dissertation is devoted to discussing different aspects of linguistics as an empirical science. The study includes different data sources and discusses what a linguist can conclude on the basis of such data. One central aim of the thesis is to present a selection of methodological options, including both data collection and data analysis (i.e. advanced statistical methods), available for studying grammatical synonymy.

The aim of the study is to present a selection of semantic and morphosyntactic variables that have an effect on the alternation between the synthetic locative case construction and the corresponding analytic adpositional construction in Estonian. Although the dissertation focuses on the Estonian adessive and the adposition *peal* 'on', the general methodology and results are applicable to alternations between other locative cases and adpositional constructions in Estonian and related alternation phenomena. The focus is on methodological pluralism – both corpus based studies as well as experimental research is employed. Emphasis is placed on the quantitative analysis of the

data, since it is believed that theoretical discussion gains in strength when it is based on empirical research.

The dissertation comprises six chapters.

The first chapter introduces the theoretical assumptions of the thesis. It gives an overview of grammatical synonymy and discusses some of the key aspects of usage-based linguistics and the related notions: frequency, Cognitive Grammar and the concept of construal, Construction Grammar and the concept of construction. The second part of the first chapter gives a lengthier treatment of linguistics as an empirical science. Since emphasis is on methodology and the nature of linguistic data, three specific data sources are considered: intuition and introspection, corpus linguistics, and experimental linguistics. The chapter ends by stressing the importance of methodological pluralism.

The second chapter describes in detail the specific linguistic phenomenon under study. It gives an account of the Estonian adessive case construction and the adpositional construction with *peal* 'on'. In addition, previous research on similar topics in Estonian and other languages is discussed. A more comprehensive overview is provided of the various alternation phenomena that have been studied in detail (e.g. the English genitive and dative alternation, and particle placement) and which serve as an inspiration for the present study as well. The third part of the chapter introduces the specific linguistic variables included in the studies. It provides both reasons for selecting these variables and the relevant examples. The chapter ends with listing the objectives and the general predictions of the studies reported in the dissertation.

The third chapter is a lengthier discussion on the selection and implementation of methodology and the statistical techniques. Since emphasis is placed on empirical research, description of the exact methodological designs and the specific methods is crucial. The first part of the chapter focuses on describing the two corpus studies and the two rating tasks. Some practical considerations for employing these methods are also discussed. The second part of the chapter is devoted to the statistical techniques. Both univariate as well as multivariate techniques are discussed. Although the primary focus is on linguistics, a fair amount of technical descriptions are provided in order for the reader to comprehend the necessity and essence of employing the statistical techniques. A fairly detailed introduction to these techniques is necessary for comprehending the actual analysis of the data as well interpreting the results.

Chapters 4 and 5 report the results of corpus and experimental studies respectively. In addition to presenting both the univariate as multivariate results, these two chapters provide summaries of the main results, although a more detailed discussion is left for Chapter 6. The sixth chapter compares the results obtained by corpus and experimental studies. Both converging as well diverging evidence is reported. In addition, the results of the present studies are compared to the results of previous research. The chapter ends with an assessment of the methodologies for future research on Estonian locative cases and adpositions and the related phenomena.

1. THEORETICAL ASSUMPTIONS

This chapter gives an overview of the major theoretical assumptions underlying this study. It starts with a definition of grammatical synonymy and discusses the notion of near-synonymy. Given that the present study conforms to the assumptions of a usage-based approach to the theory and description of language, section 1.2 highlights some of the relevant aspects of such an approach. Two specific approaches are especially pertinent – Cognitive Grammar and Construction Grammar. In addition, some key terminology used in the thesis is presented, zooming in on the notions of construal, construction, and frequency. The second part of this introductory chapter focuses on the discussion of linguistics as an empirical science. It gives an overview of the aspects of an empirical science and comments on the current situation within cognitive linguistics. Three specific data sources are looked at in detail since all of these are employed in the present study as well – intuition and introspection, corpus linguistics, and experimental linguistics. The chapter ends by stressing the importance of methodological pluralism.

1.1 Grammatical synonymy

Similarly to Divjak's (2010) study on near-synonymy in the Russian verbal domain, this work proceeds from the perspective of onomasiological variation[2]. It looks at two alternative linguistic means used to designate the "same" concept or linguistic function. In Langacker's (1987, 2008) terminology, it looks at two construals that profile the same relationship. Differently from Divjak's (2010) work, however, this thesis is not a study on lexical synonymy, but what is here termed **grammatical synonymy**. In semantics, much attention is devoted to lexical synonymy and no definition is given for grammatical synonymy. As a working definition, grammatical synonymy is defined in the dissertation as synonymy pertaining to the conceptual "sameness" of two grammatical constructions, e.g. a locative case and a locative adposition. In line with two lexical words, two grammatical constructions are treated as synonymous when the semantic similarities are more salient than the differences, and when they do not contrast with each other (Cruse 2004: 154–157).

Taken in its most general sense, synonymy refers to sameness of meaning. However, such a conception is neither particularly helpful nor necessarily valid. Absolute synonymy seems functionally unmotivated, and yet synonymy is frequent, be it at the level of lexicon or grammar. The question of synonymy opens up interesting questions: how different or similar are the specific linguistic units; if different, where are the differences located, and are the

[2] The study of polysemy and near-synonymy in cognitive linguistics can be seen as a reworking of the Structuralist semasiological – onomasiological distinction, where polysemy relates to semasiological variation and synonymy to onomasiological variation (see Geeraerts, Grondelaers, and Bakema 1994).

differences permissible enough so as not to destroy synonymy. All of these questions will be addressed in the present study by looking at the alternation between the adessive case and the adposition *peal* 'on' to express a support-relation in Estonian. By employing a range of different techniques, I reach the following conclusion: although the meanings of the two constructions are relatively close, there are certain semantic and morphosyntactic variables that play a role in determining which construction is actually used by a language speaker.

Although cognitive linguistics is well suited to deal with a linguistic phenomena such as synonymy (Divjak 2010: 5), the amount of cognitive linguistic work focusing on the issues of polysemy rather than synonymy is considerably larger. However, there is no empirical validation to such a claim since no statistical study exists contrasting the number of the relevant studies. Moreover, Glynn (2010: 90) points out that "the study of lexical synonymy has a long tradition within Cognitive Linguistics", making reference to such pioneering work as Lehrer (1982), Schmid (1993), Geeraerts et al. (1994), and Rudzka-Ostyn (1995). The current cognitive linguistic studies on issues related to synonymy can broadly be divided into two groups: those that study lexical near-synonymy (e.g. Divjak and Gries 2006, Arppe and Järvikivi 2007, Divjak 2006, Arppe 2008, Divjak 2010), and those that study syntactic alternations (Dąbrowska 1998; Gries 1999, 2001, 2002, 2003a, 2003b; Rosenbach 2003; Bresnan et al. 2007; Hinrichs and Szmrecsanyi 2007; Szmrecsanyi and Hinrichs 2008; Bresnan and Ford 2010; Szmrecsanyi 2010, in press a, in press b; Shih et al. to appear; Wolk et al. to appear). The present study falls into the second group as it looks at two alternative grammatical constructions – that of the synthetic case construction and the analytic adpositional construction. However, differently from the vast majority of the alternation studies cited in this dissertation, it does not study a word-order alternation such as the alternation between the *s*- and *of*-genitive, particle placement or dative shift in English, and most importantly, it looks at a language other than English.

The following two trends within cognitive linguistics are especially well suited for the study of synonymy: Cognitive Grammar and Construction Grammar. In the following section I will elaborate on some of the key notions in these fields that are important from the perspective of the current study.

1.2 Usage-based linguistics

The present study advocates a usage-based perspective on language (see, inter alia, Langacker 1987, Barlow and Kemmer 2000).[3] Naturally, there is no one

[3] While Langacker's Cognitive Grammar (1987, 2008) is generally considered a usage-based model par excellence, opinions differ as to what extent Construction Grammar is considered a usage-based model. Croft and Cruse (2004: 291 Ftn.1) mention that the latter is intended as a competence model instead. However, linguists working within the framework of Construction Grammar put a similar emphasis on the importance of actual language use

single usage-based model. Rather, we can talk about a range of theoretical and empirical approaches to the description of human language, which despite numerous differences in methodology share a number of fundamental assumptions (Kemmer and Barlow 2000: vii–xxi). Specifically the following four assumptions hold for the present dissertation:

1. There is an intimate relation between linguistic structures and instances of language use.
2. Frequency of instance is a prime factor in the structure and operation of language.
3. Usage data is important in theory construction and description.
4. The linguistic system is interconnected with non-linguistic cognitive systems.

Croft and Cruse (2004: 292) point out two usage-based properties of grammatical units in particular that are assumed to affect grammatical representation in a speaker's mind: the frequency of occurrence and the meaning of the words and constructions in use. I will first discuss some of the aspects of frequency and then move on to discussing meaning and other relevant notions, e.g. construal and construction.

Before moving on, however, another relevant aspect about the usage-based model needs to be pointed out – usage-based models contrast with the traditional structuralist and generative models of grammatical representation in that they are nonderivational models and eschew the need for any kind of transformation or derivation. Two sentences generally analyzed as being transformationally related, e.g. the English Dative Shift, are instead claimed to be semantically distinct by virtue of their different grammatical organization (Langacker 1987: 39). Furthermore, in the usage-based model, properties of the use of utterances in communication and the structure of the grammatical form determine the representation of grammatical units in a speaker's mind; in the structuralist and generative models, only the latter determines their representation in the mind (Croft and Cruse 2004: 292).

Frequency. It is indisputable that frequency plays a major role in any usage-based account of linguistic phenomena. Any piece of linguistic research that aims to be empirically valid and pay full tribute to the commitment of looking at actual language use involves counting some type of observed frequencies of the linguistic structure under study. It is equally indisputable that the frequency of structures has an effect on how these structures are represented in the human mind. For example, it is claimed that higher frequency of a unit or pattern results in a greater degree of entrenchment, i.e. cognitive routinization, which

---

and share other assumptions characteristic of a usage-based approach. From the perspective of the present study, the two cognitive linguistic trends are therefore both taken to belong to the domain of usage-based linguistics.

affects the processing of the unit (Kemmer and Barlow 2000: x). In cognitive linguistics, the notion of ***entrenchment*** and its correlation with frequency goes back to Langacker:

> "Linguistic structures are more realistically conceived as falling along a continuous scale of entrenchment in cognitive organization. Every use of a structure has a positive impact on its degree of entrenchment, whereas extended periods of disuse have a negative impact. With repeated use, a novel structure becomes progressively entrenched, to the point of becoming a unit; moreover, units are variably entrenched depending on the frequency of their occurrence." (Langacker 1987: 59)

Entrenchment should be kept apart from the notion of ***salience***[4] since the latter does not refer to the degree of routinization, but rather to temporary activation (Schmid 2010: 127 Ftn. 8). It also makes sense to keep entrenchment separate from the notion of ***conventionalization***: the former is a matter of individual minds and the latter pertains to speech communities (Schmid (2010: 127 Ftn.10). The two processes are, of course, intertwined and the "crucial link [...] is frequency of use and exposure, which on the one hand reflects degrees of conventionalization in the speech community and on the other hand enhances entrenchment in individual minds" (Schmid 2010: 117).

There exists a vast body of (psycho)linguistic research that suggests that frequency is one of the major factors for the ease and speed of lexical access and retrieval (for an overview, see Ellis 2002). There is a general frequency effect which means that the recognition and processing of linguistic structures is related to their frequency of occurrence (Ellis 2002: 152). For example, Bybee and Thompson (1997) demonstrate how three frequency effects shape grammar. The first two effects, referred to as the Reduction Effect and the Conserving Effect, relate to **token frequency**; the third effect relates to **type frequency**[5]. The two first effects of high token frequency are very different in nature and appear to condition somewhat different results. High token frequency both promotes change and renders constructions resistant to change (Bybee and Thompson 1997: 378). The second aspect is related to Langacker's (1987) notion of entrenchment – the more a form is used, the more its representation is strengthened (Bybee 1985). Type frequency, on the other hand, plays a role in productivity, i.e. that a linguistic pattern will apply to new forms (Bybee and Thompson 1997: 384). Again, there are implications for the underlying cognitive mechanisms. It is claimed that high type frequency ensures that a construction will be used frequently, which in turn strengthens its representation schema (*ibid.*). As attractive as such claims are, there are some serious caveats

[4] In his 2008 book, Langacker considers *salience* interchangeable with *prominence* (Langacker 2008: 66).

[5] Token frequency is the count of particular words or specific phrases in text; type frequency counts how many different lexical items occur in a certain pattern or construction (Bybee and Thompson 1997: 378).

here as pointed out by, for example, Arppe and Järvikivi (2007), Schmid (2010), and Divjak (2008, in progress). Many researchers have started to question the previously assumed fairly direct correlation between observed frequencies of linguistic units and the degree of entrenchment and salience of these units in the minds of speakers. The crucial point Divjak (2008) and Schmid (2010) are making is that there is still a long way to go in understanding the nature of frequency and its relation to cognitive underpinnings such as entrenchment. Linguists should take more care when interpreting quantitative observations in cognitive terms.

As the present thesis places emphasis on empiricism, it cannot ignore the role of frequency. Both type and token frequency of the adessive case construction and the adpositional construction with *peal* 'on' are looked at (see Chapter 4, especially distinctive collexeme analysis). From the perspective of the current study, the following distinction between types of frequency is also relevant: **absolute** vs. **relative frequency**. Divjak (2008: 214) reports that it has been shown in a number of studies that relative frequency outperforms raw or absolute frequency. Opinions differ, however, as to which type of relative frequency performs best. There are at least three different ways how to approach relative frequency: attraction (Schmid 2010), reliance (Schmid 2010), and collostructional strength (Stefanowitsch and Gries 2003). For example, in relation to the adessive case construction, attraction refers to the relative frequency of tokens of nouns that occur in the adessive construction, while reliance refers to the relative occurrence of the same noun in other constructions besides the adessive construction. Collostructional strength (Stefanowitsch and Gries 2003) is a mathematically more sophisticated method for calculating the relative frequencies of linguistic structures. It relies on statistical tests of significance between the observed and expected frequencies and also takes absolute frequency into consideration. However, although this method is mathematically more sophisticated than Schmid's (2010) calculations of attraction and reliance, collostructional strength has its own weaknesses and is open to criticism (cf. Schmid 2010, Baayen 2011)[6]. The conclusion to be drawn is that frequency of occurrence is a much less objective measure than most researchers of quantitative cognitive linguistics assume. It involves important theoretical decisions such as what to retrieve and count as valid tokens, and how to calculate frequency.

Another relevant aspect of frequency from the perspective of the present study pertains to the relationship between frequency counts and acceptability judgements since both types of data are exploited here. For example, Kempen and Harbusch (2005), Arppe and Järvikivi (2007), Divjak (2008), Bermel and Knittl (in press) have directly explored the relationship between frequency of occurrence in a corpus and the degree of acceptability in an acceptability judgement task. All studies reach a similar conclusion – there is a discrepancy between the frequency counts and the grammaticality ratings (Kempen and

[6] See Chapter 3, section 3.1.2 of the thesis for a more detailed discussion of this method.

Harbusch 2005: 330). Whereas a relatively higher frequency correlates with acceptability, relatively lower frequency does not and can be judged either acceptable or unacceptable (Arppe and Järvikivi 2007: 151, Divjak 2008: 213). In other words, frequency entails acceptability and unacceptability entails infrequency; acceptability, on the other hand, can entail either frequency or infrequency. As the results of the present study also demonstrate – a construction with lower relative frequency in a written corpus can receive a higher rating in a meta-linguistic judgement task (see Chapter 5). Therefore, there is converging evidence for Featherston's (2005a: 204) arguments that "frequencies and well-formedness judgement correlate with the "best" structures, but provide no information about "poorer" candidates".

Cognitive Grammar and the concept of construal. One of the most important theoretical starting points of the present thesis is that grammar is meaningful; the elements of grammar are similar to vocabulary items in that they have meanings in their own right (Langacker 2008: 3). Cognitive linguistics rejects the objectivist account of meaning (Lakoff 1987, Johnson 1987). The latter implies that meaning can be described in terms of objective language-world relationships. Instead, language is a representation of underlying conceptual structures and processes, which are grounded in the human body and in our experience of the world around us. From the perspective of Cognitive Grammar, and cognitive linguistics more generally, meaning involves both the conceptual content and how the content is **construed**:

> "The full conceptual or semantic value of a conceived situation is a function of not only its content (to the extent that one can speak of content apart from construal), but also how we structure this content with respect to such matters as attention, selection, figure/ground organization, viewpoint, and level of schematicity." (Langacker 1987: 138)

The fact that in language we have alternative linguistic ways (e.g. the locative cases and adpositions in Estonian) to express similar conceptual content allows for a language user to choose among a variety of grammatical and lexical items to construe an experience or a situation. These different ways, in turn, convey different conceptualizations. It is thus the hypothesis of cognitive linguistics that semantics is conceptualization and not purely truth-conditional (Croft and Cruse 2004: 40).

Any sentence involves a multitude of construals of the experience and everything that makes up the grammatical structure of a sentence involves conceptualization. However, the questions about the relationship between language, thought and experience are not self-evident. Langacker (2008: 85) is of the opinion that this conceptual view of meaning does not necessarily imply that semantic structure is directly accessible to intuition or introspection. In normal, unreflective language use we are quite oblivious to construal – we are interested in what is being said, not the specific ways in which we construe the content.

This view is echoed by Lakoff who claims that "experience does not *determine* conceptual systems, but only *motivates* them" (Lakoff 1987: 310). It is only through linguistic analysis and by seeking converging evidence that a characterization of semantic structure is achievable (Langacker 2008: 85).

Construction Grammar and the concept of construction. The cognitive linguistic approach to syntax is generally referred to as Construction Grammar where grammatical knowledge is represented as constructions, i.e. symbolic units of form and meaning pairings (Croft and Cruse 2004: 257). The field of construction grammar boasts different versions of it: Kay and Fillmore's Construction Grammar (Kay and Fillmore 1999), including Frame Semantics (Fillmore 1982); the construction grammar of Goldberg (1995, 2006) with a focus on argument structure constructions; and Croft's Radical Construction Grammar (Croft 2001). According to the overview of construction grammars by Croft and Cruse (2004: 257–290), the different versions all conform to the following three key principles: the independent existence of constructions as symbolic units, the uniform representation of grammatical structures, and the taxonomic organizations of these structures in a grammar. What vary are the terminology used and the means for description and representation of constructions.

Croft and Cruse (2004: 257) also consider Langacker's Cognitive Grammar (1987, 2008) as a variant of construction grammar. It is true that as a cognitive linguistic approach to syntax it clearly fits into this list, but in the present thesis it is considered separately. The reason for doing this comes from Langacker himself (2005), who considers Construction Grammar and Cognitive Grammar as two separate theories that have developed alongside each other, but reached similar theoretical conclusions independently. Despite listing a number of similarities between the two approaches (Langacker 2005: 158), he also stresses some of the differences (Langacker 2005: 159–162). For example, Langacker points out that differently from Construction Grammar, the idea of generativity is not present in Cognitive Grammar (Langacker 2005: 159). Similarly to Langacker (2005: 158), the present thesis considers Construction Grammar and Cognitive Grammar to be two independent theories of grammar which share many theoretical premises.

This dissertation does not subscribe to any specific version of Construction Grammar. The majority of the different versions listed above have mainly focused on verbs and verbal semantics. The topic of this study is, however, locative cases and adpositions. Hence, although the very general principles that underlie all types of Construction Grammar apply to this dissertation as well, no constructional schemas are presented for the Estonian adessive and the adposition *peal* 'on'. This remains a task for the future, since such an account requires a full acknowledgement of the polysemy of the two constructions – something that is not the focus here. Instead the very general notion of **construction** is made use of. Grammatical constructions consist of pairings of form and meaning that are fundamentally symbolic units. Generally speaking,

constructions are non-derived patterns, abstract expressions at the schematic level of language featuring slots to be filled with elements at a more specific level (Divjak 2010: 7). The **adessive construction** consists of a noun or a pronoun inflected for the adessive case that expresses a support-relation. The ***peal*-construction** consists of a noun or a pronoun inflected for the genitive case followed by the postposition *peal* that expresses a support-relation. Although both constructions can be used to express a situation where one object is placed on top of another object, it is believed that a difference in syntactic form implies a difference in meaning (e.g. Bolinger 1968, 1977); cf. Goldberg's *Principle of No Synonymy* (2005: 67): "If two constructions are syntactically distinct, they must be semantically or pragmatically distinct".

1.3 Linguistics as an empirical science

The aim of this section is to give a short overview of linguistics as an empirical science. In addition to providing a short historical overview, it assesses the current situation (especially within cognitive linguistics), and highlights the key aspects of doing empirical research. One of the aims of this section is also to discuss the nature of linguistic data and what kind of constraints on methodology the nature of the data places. One of the major constraints of doing empirical linguistic research is the kind of research questions linguists ask. Cognitive linguists are interested in researching meaning, but meaning is a particularly difficult subject to study empirically. In particular, three types of methodologies and the type of evidence they provide are discussed: corpus linguistics, experimental linguistics, and intuition-based evidence. The section ends with the conclusion that other researchers have drawn – the best result is obtained by combining different methodologies and sources of data.

1.3.1 The current situation in cognitive linguistics

As with any empirical science, data are central to linguistics. However, as Gilquin and Gries (2009: 1–2) stress, there is no agreement on what exactly qualifies as data. There is even less agreement as to how linguistic data should be obtained, analysed, evaluated and interpreted. Although the aim here is not to provide a thesis-length discussion of the different types of data and different methodologies, it is considered important to highlight the specific advantages and disadvantages of different methods since linguistic empiricism holds a central position in this study. It is also important to realize the differences between these various methods and especially the kind of linguistic evidence they produce. While Labov in his 1972 article states that "the status of *methodology* has fallen so fast and so far that it now lies in that outer, extra-linguistic darkness where we have cast speculation on the origin of language and articles about slang" (Labov 1972: 97), the recent trend in linguistics has been quite the opposite and the field in general is in danger of becoming obsessed with

empiricism (cf. Schmid 2010). With the advances made in both data collection as well as data analysis, the concept of "proper" methodology in linguistics has become a widely debated and discussed topic. As proof of this, one only needs to check recent bibliographies on linguistic methodology as well as the electronic mailing lists for linguists for calls for different workshops on empirical methods (e.g. the EMCL[7] series). This shift in paradigm has resulted in the publication of various edited volumes and monographs on linguistic methodology, e.g. Kepser and Reis (2005), Gries and Stefanowitsch (2006), Gonzalez-Marquez *et al.* (2007), Glynn and Fischer (2010), Newman and Rice (2010), Glynn and Robinson (*to appear*), Sharma and Podesva (*to appear*).

As a reaction to the extreme empiricism of the behaviourists, the 1950s and 1960s saw the rise of introspection as the main source of evidence in linguistics, especially within the domain of formal syntax. The heydays of introspection and informally collected linguistic acceptability judgements went hand in hand with those of transformational-generative grammar. A much voiced concern about this kind of approach is that data which conflicted with a linguist's own judgements were simply discarded (Labov 1972: 106, Gilquin and Gries 2009: 2). However, the mid-1990s witnessed a shift in paradigm (Kepser and Reis 2005: 2) and the word *methodology* has ceased to be a taboo word in linguistics. It is no longer "an open secret that the rough and ready exploitation of grammatical intuitions has run its course from exhilaration to despair" (Labov 1972: 97), but a generally accepted status quo. Naturally, no self-respecting linguist will deny the usefulness of intuitions and introspection, but it is increasingly frowned upon if one's arguments as a linguist are based solely on one's own intuition as a native speaker of the language under study. Contrary to this, it is the opinion of some that the field is about to take the other extreme with voiced concerns about "empirical imperialism" (cf. Geeraerts 2006: 34, Schmid 2010). Although there is talk about the "empirical revolution" in linguistics (Geeraerts 2006), a cursory look at the current situation within the linguistic framework most relevant for the present thesis – that of cognitive linguistics – has left researchers with a fairly grim outlook (see, for example, the general conclusions in Sandra 1998, Geeraerts 2006). It seems there is still a long way to go before doing proper empirical research will be a feather in the cap of cognitive linguists. This is especially evident when we compare the situation in linguistics with that of sociology or psychology – two disciplines that clearly have strong ties with linguistics.

Naturally, different fields within linguistics are at different stages on this road towards empiricism. As Gilquin and Gries (2009: 4) point out in their "state-of-the-art" overview of different methodologies, areas other than theoretical syntax and semantics are doing relatively well as pertains to methodological diversity, listing phonetics, sociolinguistics, psycholinguistics, and cor-

[7] EMCL stands for *Empirical Methods in Cognitive Linguistics*. It is "a series of interdisciplinary workshops deisgned to help students and researchers acquire the empirical skills necessary to test Cognitive Linguistic hypotheses" (http://www.cogling.org/emclconfs.shtml, accessed 5 April 2012).

pus linguistics. We may also add second language learning, language acquisition, sign language and gesture research to this list (Geeraerts 2006: 35). Sociolinguistics and psycholinguistics are in an especially favourable position – both sociology and psychology as scientific disciplines have for a very long time had a clear idea of what "methodology" is and fully acknowledge the importance of empirical research. It is fair to say that there exists at least some sort of consensus on how to go about collecting evidence and there is general agreement on the necessity of adequate data analyses techniques. In comparison, there is no common empirical ground in linguistics in the form of a set of data derived through a generally accepted method (cf. the role of experimentation in psychology, for example) (Geeraerts 2006: 26). Of course the nature of linguistic data does not allow us to produce similar empirical research as undertaken in the sciences[8], linguistics as a discipline still has a lot to learn from disciplines that standardly do employ empirical methods (e.g. sociology, psychology, neurosciences), regardless of the fact that the data in these disciplines pose the same kind of difficulties as language does. The object of study in all of these is the human being. Human behaviour is rarely determined by a single cause and requires the use of adequate methodology to cope with the complexities of such phenomena as everyday language use (Gries 2003a: 172).

In Geeraerts' (2006: 31) view, cognitive linguistics is in a particularly advantageous position for leading the methodological progress of linguistics by virtue of its theoretical assumptions: its cognitive nature, its usage-based perspective, and its contextualizing approach. It is especially beneficial to study such theoretical assumptions using experimental data, corpus material, survey techniques, and advanced quantitative analysis. However, there is a paradox here: one dominant feature of cognitive linguistics is the emphasis on the analysis of linguistic meaning over linguistic form (Geeraerts 2006: 28). On the one hand, cognitive linguists emphasise the importance of employing different methodologies and following the more rigorous empirical research protocol of, for example, sociology and psychology, but on the other hand, the quintessential question a cognitive linguist ends up asking concerns linguistic meaning. Meaning, however, is something that does not exhibit itself in the corpus data, but in the human mind and not in language per se (at least in the opinion of cognitive linguists), i.e. it is something that is fairly difficult to measure. This seems to support the view that introspection is the only reliable method of direct access to semantic phenomena. Therefore, although scientific objectivisism is a desirable ultimate goal, it can never become a reality – there will always be some subjectivity at some stage of cognitive linguistic research as is the case in other disciplines as well. Still, employing a more rigorous methodology helps to diminish the amount of subjectivity and the use of adequate data analysis

[8] In a discipline like linguistics we are always talking about a degree to which a linguist is right or wrong – the right or wrongness of a particular analysis is not as obvious as in the natural sciences (Sandra 1998: 364) and the precise quantitative predictions that are possible in physics may not be possible in linguistics (Langacker 1987: 33).

methods (i.e. advanced statistical techniques) will ensure at least some advancement in objectivity.

Cognitive linguistics is already witnessing considerable progress: psycholinguistic work in the form of elicitation and experimentation has long been part of cognitive linguistics (Gibbs 1994, Sandra and Rice 1995 referred to in Geeraerts 2006: 32)[9] as has been corpus data. Geeraerts (2006: 33) reports an increasing trend in the articles and books cited in the Cognitive Linguistic Bibliography[10] between the years 1985–2004 that refer to empirical approaches. He concludes that there is an intensification of the appeal of empirical methods within cognitive linguistics and the growing importance the cognitive linguistics community attaches to empirical methodology. This is a considerable improvement compared to earlier work, which according to Geeraerts (2006) is only good for interpreting phenomena that need further empirical support. Geeraerts (2006: 41) offers fair criticism in claiming that the work of Langacker[11], Talmy and Fauconnier is in the manner of highly traditional linguistics with emphasis on theory-building rather than data gathering and data analysis. What the field needs are operationalizations of such interpretative hypotheses of the "founding fathers" that allow for hypothesis-testing. But the fact of the matter is that it is notoriously difficult to operationalize such cognitive terms as salience, entrenchment, prototypicality, etc. However, cognitive linguists should not be defeatists and give up on studying these issues, but use their ingenuity instead and come up with testable hypotheses – it should be possible to validate and replicate results. Only once we have made the systematic attempt at operationalizing linguistic hypotheses about theoretical concepts and statements can we talk about an "empirical revolution in linguistics" (Geeraerts 2006: 27).

At the same time, Langacker (1987: 33) quite rightly points out that "no one body of methodological assumptions can be considered definite" and that "every methodology has its limitations" (Feyerabend 1978 cited in Langacker 1987: 33). True, in our eternal quest for empiricism we should not forget that as to everything, there is a good side and a bad side. Nevertheless, researchers

---

[9] Another domain where experimental research has long been the dominant trend is spatial language research (Carlson and Hill 2007).

[10] Cognitive Linguistics Bibliography (CogBib) is a research instrument available for subscribers of the journal *Cognitive Linguistics*.

[11] If we look at what Langacker has to say about methodology, we find the following statements:

> "I would greatly prefer not having to discuss methodological issues, for a number of reasons. First, I profess little expertise in methodological concerns or the philosophy of science. A second reason is an oftnoted tendency for the amount of attention an author devotes to methodological questions to correlate inversely with the extent of his actual descriptive contribution. A third factor is the consistently inconclusive nature of methodological disputes, which remind me very much of legal disputes." (Langacker 1987: 31)

Yet, at the same time, Langacker (1987: 45) does acknowledge the conception of linguistics as an empirical discipline.

should not be put off if different methodological principles lead to conflicting results. It is not so much converging evidence as diverging evidence that leads to the improvement of scientific theories. In the words of Labov: "methodology is careful and conscientious search for error in one's own work" (1972: 99).

Geeraerts (2006: 34) concludes that "the empirical approach, although on the rise, does not yet dominate the methodological landscape in Cognitive Linguistics". Much work in cognitive linguistics is still based on what Geeraerts (2006: 35) has termed the "traditional methodology of conceptual analysis", lacking in the explicit attention for empirical data. In addition to simply stating the grim reality, Geeraerts (2006: 39–44) highlights some of the reasons for this status quo, pinpointing two reasons: practical reasons and reasons of principle. Neither set of reasons are compelling enough to be used as an excuse for not doing empirical research, but they are, nevertheless, something that needs to be taken into account. While quantitative and qualitative data analysis and training in experimental techniques is part and parcel of the education and training in psychology and sociology, such an education is not (yet) a standard part of curricula in linguistics. As Geeraerts puts it, the "practical steps to be taken are indeed numerous and often cumbersome" (2006: 37) and didactic efforts are necessary to make cognitive linguists acquainted with different data collection and analysis methods. Recent years have seen the publication of some very useful practical sources (Baayen 2008, Gries 2009a, 2009b), but it remains the responsibility of individual researcher to acquaint themselves with such methods rather than something that can be taken for granted. Already in the 1970s Labov suggested that "a lack of professional orientation towards equipment has been a serious impediment in the development of the study of language in everyday life" (Labov 1972: 110). The situation nowadays is very similar with researchers struggling to use statistical packages, scanners, MRI, eye-tracking devices, and other such technical equipment. Of course the important question is do we, linguists, need to be able to use the same kind of technology as the psychologists do – where do we draw the line between a linguist and a psychologist; I return to this issue below.

The second set of reasons – reasons of principle – relate back to the paradox that inhibits cognitive linguistics: "is an objectivist methodology compatible with a non-objectivist theory?" (Geeraerts 2006: 42). Geeraerts (2006: 43) entertains two sets of arguments, the minimalist and maximalist reading of this question. The first implies that empirical evidence is not sufficient for a proper study of meaning and the second that it is not necessary. According to Geeraerts (2006: 43) both arguments can easily be countered. If we are to make progress by achieving a higher degree of comparability between different theoretical approaches, clinging to the traditional, highly subjective methods of research will not help in achieving this goal. In order to be able to compare theories, we first need a common empirical ground, i.e. data with regard to which to compare different theories (Geeraerts 2006: 26).

As an example, let us consider one very controversial issue in lexical semantics – that of polysemy; more specifically, the status of polysemy net-

works in cognitive linguistics. Although a highly appealing concept, lack of explicit criteria for distinguishing between usages has led different linguists to put forward different networks for one and the same preposition (Cuyckens et al. 1997: 36). In general, then, it is the analysis skill and subjectivity of individual researchers that determine the content and format of specific network models and there seems to be no way to test the validity of one network over the other. Yet another set of questions pertains to how these network analyses relate to the mental structures of the language user's mind (see, for example, the discussion in Sandra and Rice 1995, Cuyckens et al. 1997, Croft 1998, Sandra 1998, Tuggy 1999, Taylor 2006). If cognitive linguists cannot present convincing empirical evidence, comparable, for example, to the type of evidence presented by psychologists, there is virtually no way one can claim that the lexical-semantic network models of cognitive linguists are superior to the classical approaches to polysemy, save for the arguments of the kind "my theory is better than yours". As Sandra (1998: 364) has stressed, linguistics is a discipline where the risk of making the data fit one's theory is considerably greater than in other disciplines. The call for empiricism was already voiced by Sandra and Rice (1995) and Cuyckens et al. (1997: 50): in order to add validity to our claims, we need to develop the means and a will for formulating and testing specific hypotheses.

Naturally, this is not to say that every linguist should become a psychologist – it is more profitable to the discipline as a whole if every cobbler sticks to their own last. However, it is equally valid that calling oneself a cognitive linguist does not automatically mean that one is licensed to make statements pertaining to issues about general cognition (Sandra 1998: 364). Of course this raises the question "to what extent can linguists say anything about the human mind" (Sandra 1998: 362), a question that is usually followed by a list of negatives answers: they "cannot address issues pertaining to the processing of linguistic material" and they "cannot make statements about the representational format of language elements in the language user's mind either" (Sandra 1998: 375). Nevertheless, there is at least one positive answer as well: "linguists are able to make some statements on what is in the mind" (Sandra 1998: 375). Thus, in Sandra's (1998) opinion, a linguist should stick to analysing language and leave the "complicated" issues of the representational format of language elements in the mind to the hands of psycholinguists (1998: 362). According to Tuggy (1999: 364), however, this "counsel of despair is not justified": there is plenty of evidence which linguists can use to address these issues and they should not be denied the right to say anything at all unless they can immediately prove everything they think is true (Tuggy 1999: 354). True, any statements linguists come up with should be taken as mere hypotheses to be validated empirically. However, "the picture is not nearly so bleak" (Tuggy 1999: 366) and with co-operation with researchers from other disciplines who study the human (mind), e.g. psychologists, statisticians, sociologists, biologists, etc. we are able to conduct cognitive linguistic research that pays full tribute to its grand name.

Following this discussion, it is important to stress that the present work is not a thesis in psycholinguistics. The aim is to provide an adequate usage-based description of the linguistic phenomenon and to test some of the predictions and assumptions that spring from the corpus data with a set of off-line linguistic experiments. Thus, caution has been taken in positing theory-wise anything more than can be plausibly expected given the type of methodologies used to collect and analyse the data. I have only included as much theory as is licensed by the empirical data available to me. However, the concerns that empirical research automatically leads to the abandoning theory formation are unfounded (Geeraerts 2006: 45). Instead, what is at stake here is providing proof for one's theories and/or refining them accordingly. It certainly does not lead to "throwing out the baby with the bathwater" (Gilquin and Gries 2009: 3) and abandoning intuition and introspection completely. Introspection still remains an important step in the empirical cycle, albeit only one and it should be accompanied by other methods, such as corpus studies and/or linguistic experiments (Geeraerts 2006: 45).

In conclusion, it can be said that the current situation in cognitive linguistics in general is a mixed one and the crucial question raised by Geeraerts (2006: 44) and others still awaits an answer – will cognitive linguistics embrace the empirical turn or will it stay a predominantly traditional, theory-building approach? I agree with Geeraerts (2006: 27) that turning linguistics into an empirical science is not a mission impossible, but something realistically achievable. What needs to be achieved is, first of all, recognising and understanding of why this is necessary and acquiring the relevant practical know-how how to conduct empirical research, i.e. using experimental methods, survey techniques, and corpus methods. It is hoped that this dissertation brings us one step closer to achieving this aim. The following paragraphs highlight some of the aspects of doing empirical research, discuss the nature of linguistic data and provide an overview of intuition and introspection, corpus and experimental linguistics.

Aspects of empirical research. Following Geeraerts (2006: 23–24)[12] the following five features of empirical research may be identified:

1) it is data-driven;
2) it concerns the formulation of hypotheses;
3) it requires the operationalization of hypotheses;
4) it involves quantitative methods;
5) it involves an empirical cycle.

The most important of these features are the second and third. Any kind of (empirical) research starts with postulating a research question and deciding

[12] See also Popper (1959) for a more general discussion on the various steps involved in any scientific discovery.

how to study it. The trick is not just asking any questions, but expressing the question in such a manner that it is possible to verify or falsify it (cf. the fourth feature in the list). Operationalization of hypotheses is not only one of the most important aspects; it is also, at the same time, the most time-consuming. It requires "all the ingenuity of the researcher – and most of his or her time" (Geeraerts 2006: 24), because arriving at the relevant data and measurements is not an automatic process. Empirical research combines both inductive and deductive reasoning – a bottom-up approach from data to hypotheses and a top-down approach from the theoretical perspective to thinking about the data.

As to the first feature – the obsession with data, as is appropriate for any usage-based theory of language – Geeraerts (2006: 23) stresses that the more data a researcher has, the better and likewise the more sources the observations come from, the better. However, it is not necessarily true that more data will automatically lead to a more valid and reliable result. With increased data samples, the researcher is running the risk of committing what statisticians call a Type I error – rejecting the null-hypothesis when in fact the null-hypothesis is true. That is, the researcher concludes that there is a statistically significant relationship between two variables based on his or her large data sample, when this is not the case. Increase in data size considerably increases the likelihood of finding a significant result. Researchers should therefore take into account other aspects about statistical significance testing (see section 3.3.3 for details). The fourth feature involves employing the appropriate statistical techniques to determine whether observations obtained are due to chance or not. This dissertation has tried to follow this suggestion to the letter and involves a number of statistical techniques discussed in more detail in sections 3.3 and 3.4. It is my fundamental conviction that quantitative (and empirical) research is not an end in itself, but a valuable means to an end. The "number crunching" used in the dissertation is necessary to validate hypotheses and predictions made and to assess the likelihood of the results having arisen merely by chance.

Last, but not least, as natural of any scientific discipline, linguistic research involves several rounds of data collection, hypothesis testing and interpretation of the results (Geeraerts 2006: 24). The results of this study should be therefore seen as the first round of the empirical cycle with hope for future verification or falsification.

The nature of linguistic data. Table 1 presents different kinds of data linguists have at their disposal. The table is adapted from Gilquin and Gries (2009: 5) and the different types of data are listed according to the naturalness of production/collection. Such a presentation of data types on a scale of naturalness is clearly an over-simplification as to classifying the different activities on a unidimensional scale. Arppe, for example, argues that "making linguistic judgements is just as natural a linguistic activity as language production, albeit of a different quality" (Arppe et al. 2010: 18). The former figure at the lower end of the scale in Table 1, and the latter at the top end.

Table 1. Kinds of linguistic data (based on Gilquin and Gries 2009: 5)

	Data source
more natural ↓	1. Corpora with written texts (e.g. newspapers, weblogs)
	2. Example collection
	3. Corpora of recorded spoken language
	4. Data from interviews
	5. Experimentation requiring subjects to do something with the language they usually do anyway:
	- sentence production as in answering questions
	- picture description
	- *etc*.
	6. Elicited data from fieldwork (e.g. responses to "how do you say X in your language?")
less natural ↓	7. Experimentation requiring subjects to do something with the language they usually do not do:
	- sentence sorting
	- measurements of reaction times in lexical decision tasks
	- word associations
	- acceptability/grammaticality judgements
	- *etc.*

Some of the most important points that Table 1 tries to make are the following: first, there is no discrete boundary between a corpus and an experiment and we are instead dealing with a continuum; second, linguistic data differs as to its naturalness. Clearly, looking at a corpus of written texts gives access to a more natural and spontaneous language use than experimentation. Linguists often make a distinction between observational and experimental methods (Gilquin and Gries 2009: 8). As a comparison to the list in Table 1, Labov (1972) lists the following types of data: texts, elicitations, intuitions, experiments, and observations; we would find linguists employing such activities "in the *library*, the *bush*, the *closet*, the *laboratory*, and the *street*" respectively (Labov 1972: 99). In the present thesis, both ends of the continuum are presented. I will discuss both corpus data from written texts (section 3.1.) and data derived from experimentation that requires the subjects to provide judgements (section 3.2.).

As Table 1 indicates, it is not always easy to determine whether something is an experiment or an observation. Geeraerts (2006: 36), for example, points out that especially with techniques like production tasks, the borderline between elicitation and observation is fuzzy. A production task may very well be considered as a method of corpus elicitation. The way the data are collected resembles an experimental design, but the nature of the data to be later analysed is that of a corpus of elicited observations.

Clearly, linguists differ as to what they consider the "best" type of data. Essentially, it all boils down to the specific research question and topic. Labov

(1972: 97), who is a sociolinguist, is of the opinion that the best data a linguist can have are observations in the street, i.e. the every-day language as spoken by people in their natural surroundings. This kind of data should be the most systematic basis for linguistic theory, but as Labov (*ibid*.) himself concedes, it is the most difficult to obtain. There exists something referred to as the observer's paradox: "[t]o obtain the data most important for linguistic theory, we have to observe how people speak when they are not being observed" (Labov 1972: 111). Once you start conducting surveys, interviews or recording spoken speech, you already lose in naturalness. Written texts, elicitations and intuitions are much more easily accessible than observations (Labov 1972: 105). Some even believe that the full structure of language can only be witnessed in its most literary developments; speech is considered somewhat inferior, because it is incoherent (cf. Labov 1972: 109). At the same time, there remains room for introspection and intuition. Many linguistic phenomena are too rare and infrequent[13] which leads to the complementation of observations with intuitions and elicited data (Labov 1972: 117).

Gries (2002) and Geeraerts (2006: 27) consider observational corpus analysis as a method that is most typically linguistic, while experimental techniques are shared with psychology and surveys and field techniques with sociology. The natural origins of corpus data make corpus materials the primary empirical source for a usage-based approach to linguistics. In contrast to Gries (2002) and Geeraerts (2006) the preferred data sources for Featherston (2005a) in syntactic research are acceptability or grammaticality judgements. Featherston's (2005a) arguments suggest that judgements yield data on all linguistic structures irrespective of their frequency of occurrence in a corpus. I do not consider one set of data as supreme to another.[14] From the perspective of the research question it seems necessary to exploit all three main types of data – data based on introspection, corpora and experiments. It is, however, vital to bear in mind that each set of data comes with its own advantages and disadvantages. It is important to acknowledge these aspects and take them into account before drawing any far-reaching theoretical conclusions about the linguistic phenomenon under study. The next sections will therefore take a more detailed look at the essence of the three data sources and will conclude with an emphasis on combining different methodologies, i.e. methodological pluralism.

[13] Cf. The Zipfian (Zipf 1935) distribution that characterises linguistic data – there are a few very common linguistic phenomena, a middling number of medium frequency phenomena, and many low frequency phenomena (Manning and Schütze 1999: 24). Corpora may provide insufficient data as to the last type of linguistic phenomena and it may be necessary to resort to other types of data.

[14] It is important to point out here that all linguistic evidence is subject to the same constraint – it is much easier to prove that something exists than to prove it does not exist (Tuggy 1999: 360). Tuggy emphasises that "clear proof of the presence of a semantic connection or distinction is pretty much proof of its presence: correspondingly strong proof of its absence is only proof that you did not find it" (Tuggy 1999: 360).

1.3.2 Intuition and introspection

Despite the fears of some, intuition and introspection[15] will always play a central role in linguistics. There should be no empirical imperialism and empirical research will always require conceptual analysis (Geeraerts 2006: 34). Despite being unreliable and inconsistent, introspection is widely applied as a tool in the formulation of hypotheses and interpretation of results (Arppe and Järvikivi 2007: 132). Although it seemed to Labov (1972: 97) that the "exploitation of grammatical intuition has run its course from exhilaration to despair", intuition-based studies still hold a firm position in linguistics methodology. As Labov himself concedes: "it will no doubt always be necessary to extend our observations with intuitions" (1972: 117). What should be born in mind, however, is that introspective evidence is not a superior method and if not exercised with care and evaluated carefully, it may lead to unwarranted results. However, psycholinguistic experiments not taken warily and not evaluated carefully may lead to similarly "poor" results and conclusions. As with any other method, introspective linguistic evidence should follow the criteria of empirical research outlined above.

Labov (1972: 105) discussed the notion of the *Saussurian paradox* in relation to intuition as a valid methodology in linguistics. The Saussurian paradox is that "the social aspect of language can be studied through the intuitions of any one individual, while the individual aspect can be studied only by sampling the behaviour of an entire population" (Labov 1972: 105). It is especially on issues of variation and alternation that intuitive judgements are the least regular and it is rare to find 100 per cent agreement or disagreement (Labov 1972: 107). However, modern statistical techniques do allow bringing order into chaos and there are evident benefits in using coincident intuitions that are intersubjectively valid (Cuyckens et al. 1997: 52 Ftn. 5&6, Tuggy 1999: 358).

Depending on the specific field of study and research questions, language users' intuitions may be an extremely important source of data and evidence (Tuggy 1999: 352). This is clearly the case when one studies meaning. Questions of meaning, and especially how it is represented in the mind, are not simple (Tuggy 1999: 366). Contrary to the bleak picture painted by Sandra (1998), Tuggy (*ibid.*) believes that linguistics can use many kinds of evidence, including introspection and intuition, to elucidate at least certain aspects of meaning. Since the present study is also about meaning, both my own intuition as a native speaker of Estonian as well as the intuitions of other native speakers of Estonian are fully exploited. Nevertheless, this is done in a systematic way by employing a multivariate corpus analysis and collecting the judgements of a

[15] Some researchers make a distinction between the two terms: introspective linguistic evidence and intuition (Tuggy 1999: 352). Although the former is said to cover more accurately the type of knowledge that can be gained from careful and rational consideration of what one is aware of, the two terms are used here interchangeably to refer to one and the same underlying idea.

number of subjects, followed by the use of statistical techniques to assess whether these results are obtained by chance or not.

1.3.3 Corpus linguistics

Opinions differ whether to talk about corpus linguistics as a theory or a set of methods (see, for example, Tognini-Bonelli 2001, Teubert and Krishnamurthy 2007: 1–10). From the perspective of the present study, it is not critically important to take a determined stand on this issue. Suffice it to say that corpus is regarded here, first and foremost, as a highly valuable tool for doing linguistic research. As mentioned above, for a number of researchers, corpus analysis has a number of advantages over other linguistic methods (Gries 2002, Geeraerts 2006, Arppe and Järvikivi 2007: 132). By employing corpus linguistic methods, a linguist certainly has access to natural observations of language use. In addition to the naturalness of data, another advantage of using corpora is that a considerably larger amount of data can be studied using a corpus than an experimental design (Gilquin and Gries 2009: 8). However, this should not be taken as evidence that I am asserting the supremacy of corpus data over other types of data, most notably linguistic experiments. As has been noted already above, it all depends on the specific research question and the best result is obtained by a combination of methods.

The present study takes a similar approach to corpus as Gilquin and Gries (2009: 6) who opt for a radial-category approach in defining a corpus: there are several neither necessary nor jointly sufficient criteria that define a prototypical corpus. A prototypical corpus is (Gilquin and Gries 2009: 6):

- machine readable;
- representative with regard to a particular variety/register/genre;
- balanced with regard to a particular variety/register/genre;
- has been produced in a natural communicative setting.[16]

Generally speaking, corpus studies are observational: variable levels are recorded on the basis of the observed data and not assigned before the study as in case of experiments; in addition, external validity is high (naturalness of data), but there is increased amount of noise (Gilquin and Gries 2009: 8). Gilquin and Gries (2009: 11) distinguish between the different roles a corpus may have in relation to (psycholinguistic) experimentation: the corpus may be used as a validator of the experiment, the corpus is validated by the experiment (validatee), the corpus and experimental data are used on equal footing, and the corpus may serve as a database from which experimental stimuli are composed.

Geeraerts (2006: 36) calls for using corpora not as a simple data gathering tool but as an empirical testing ground for previously posited hypotheses.

[16] The corpora used in the studies reported in the dissertation meet all of these requirements (see section 3.1 for details).

Alternatively, corpus data can be used to generate hypotheses to be validated using experimental work (cf. the corpus studies and the experimental studies reported in this dissertation). In many cases in cognitive linguistics corpora are still used as "a repository of examples that are then analysed in a traditional way" (Geeraerts 2006: 36). It is not enough to just determine whether a linguistic phenomenon is present or absent from a corpus, but a proper analysis involves the systematic exploration and statistical evaluation of the relevant frequencies[17]. According to Geeraerts (2006: 36), a usage-based linguistics needs quantification and statistical analysis, coupled with testing specific expectations or hypotheses against the corpus. In the present study, corpus and experimental data are used on equal footing, although the hypotheses tested in the rating tasks grew out of the corpus studies. In addition, by employing both the univariate (e.g. chi-squared test and t-test) and the multivariate statistical techniques (e.g. binary logistic regression)[18] in analysing the results of the corpus studies, this dissertation aims to meet the requirements of a rigorous corpus-driven study.

In addition to the obvious advantages of corpus linguistic work, one should not forget its limitations. One of the most frequent limitations pointed out in the literature is that corpora are of little help if the researcher is studying a linguistic phenomenon that is rare and does not occur with sufficient frequency in the corpus (e.g. Arppe and Järvikivi 2007: 132). If such is the case, other sources of empirical evidence have to be used. Another disadvantage is that given the size of many corpora, even the smallest differences may be significant (Gilquin and Gries 2009: 17). Therefore, corpus-linguistic work, i.e. observational data, should be complemented by experimental data in order to "separate the wheat from the chaff" (Gilquin and Gries 2009: 17).

1.3.4 Experimental linguistics

Following Cobb (1998), Maclin and Solso (2007) I consider experimental studies to be studies where explanatory variables are systematically manipulated to determine what effect they have on the response variable, i.e. the linguistic phenomenon under study. Since the subjects of linguistic experiments are human beings as in psychology and sociology, the results often come with considerable variation between subjects. As Labov (1972: 109) has pointed out – the grammar of a speech community is more regular than the behaviour of the individual. Another cause of variation in the results is the specific stimuli used and the way they are presented (so-called priming effects). Differently from a prototypical corpus where there is high external validity and low internal

[17] However, the reader is reminded about the discussion in the previous section concerning the relationships between corpus frequencies and cognitive phenomena (e.g. acceptability, entrenchement, etc.).

[18] See sections 3.3 and 3.4 for an overview of these and other statistical techniques employed in the thesis.

validity, there is a mismatch between higher internal validity and lower external validity in experiments, because normally a linguistic experiment involves artificial settings. The latter aspect, however, helps to minimise noise that permeates corpus data. The fact that experiments allow the study of infrequent linguistic phenomena and depending on the nature of the experiment, permit the study of online processes, are also listed among the advantages of experiments over corpus studies.

Nevertheless, regardless of these advantages, experimental methods are not as fully employed as one would expect given the theoretical assumptions of cognitive linguistics mentioned in the section above. As pointed out by Geeraerts (2006: 27), corpus-linguistic methods are considerably more widely used than experimental ones. One of the reasons may be that, compared to experiments, performing a systematic corpus-linguistic analysis meeting the requirements posited for a rigorous empirical research is somewhat easier. Although Geeraerts (2006: 26) reports that experimentation is the commonly accepted way to settle theoretical disputes in psycholinguistics, not any experiment will do. A proper experiment has to be adequately designed and carried out, placing thus higher practical demands on the researcher – one needs good experimental training and the ability to choose from among a set of relevant experimental designs to find an answer to a specific research question. At the same time, extensive training is also needed for doing systematic corpus linguistic research as well. I would argue that it is not only experimental research that places high practical demands on the researcher, but corpus linguistic work as well – both require extensive training which is currently not fully integrated into linguistic curricula.

Some of the relevant practical questions researchers need to ask themselves before conducting an experiment, include the following (Schütze 1996, Cowart 1997, Cobb 1998, Maclin and Solso 2007): how many participants are needed (i.e. questions pertaining to the sample size), how many test items to use per experimental condition, how to operationalize the research question (i.e. how to measure the response variable), what role do item- and task-specific effects play and how to minimize this effect, how do we know whether the hypotheses posited have been verified or falsified, what kind of instructions should be given, how long can the experiment be, how to recruit participants, and so on. Differently from corpus linguistics where practical hands-on textbooks about conducting corpus studies now do exist (e.g. Gries 2009a), there are relatively few textbooks teaching linguists how to design experiments. General textbooks and resources on how to design and analyse experiments are available (e.g. Cobb 1998, Maclin and Solso 2007 to name a few), but they offer little practical help about conducting specifically linguistic experiments. There are a few exceptions (e.g. Schütze 1996, Cowart 1997), but these tend to focus on only one type of linguistic experiments – acceptability judgements. Carlson and Hill (2007) is another excellent source on different experimental designs available for studying spatial semantics, but the authors limit themselves to discussing the general advantages and disadvantages of the different designs and offer little if

any practical guidelines how to actually set up the experiments and analyse the data.

In addition to the many practical issues, any linguist working in cognitive linguistics faces the second issue – is the experimental study of meaning possible and if yes, how. Cuyckens et al. (1997: 51) point out that the psycholinguistic study of word meaning beyond the obvious cases of homonymy presents a real challenge to the researcher. This has led linguists interested in meaning to call for an improved communication between linguists and psycholinguists who possess the relevant practical and theoretical know-how to tap into the human mind (Cuyckens et al. 1997, Sandra 1998, Croft 1998). The ultimate goal should be the combination of different methods. Swerts and van Wijk (2005), for example, were able to identify, based on corpus data, different prosodic and lexico-syntactic features coinciding with the use of a particular word order in Dutch, but needed an experimental set-up to tease these factors apart. Similarly, the present study first seeks to identify, using the corpus method, the different semantic and morphosyntactic variables that play a role in the alternation between the Estonian adessive and *peal* 'on'. As a next step, experimental methods (more specifically rating tasks) are used to look at the individual contribution of some of these variables and whether corpus results are mirrored in the rating task results.

Different types of experiments. A distinction is made between experimental techniques that are off-line and on-line. In off-line tasks, subjects are able to reflect on their own performance (e.g. when sorting or rating the stimuli), while in on-line tasks they perform a task under time-pressure with no indirect reflection on the aspects of the underlying representation structures. Another favoured tool in psycholinguistics is a reaction time experiment. All in all, there is a wide diversity of experimental techniques available to study language. Following is a non-exhaustive list (in no particular order) of some of the possible experiments used in linguistics, extracted from Cuyckens et al. (1997), Carlson and Hill (2007), Gilquin and Gries (2009): primed picture naming, primed semantic decision, sentence completion, sentence sorting task, semantic similarity judgement task, eye-tracking, self-paced reading, acceptability judgement, stimulus repetition, lexical decision task, dictation task, vocal imitation, and translation task.

Although Gilquin and Gries (2009: 13) point out that it is not unusual for psycholinguists to employ several types of experiments in one and the same study, caution should be exercised when putting two different types of experiments together. The reason, as Arppe and Järvikivi (2007: 131) stress, is that different tasks pertain to distinct linguistic processes. Again, this is not to undervalue the necessity of approaching the same research question from multiple angles, but to stress the fact that a researcher should think twice before merging, for example, the distinct experimental designs of forced-choice task and acceptability rating tasks into one experimental design. Bresnan (2007) is an example of a task where the two are merged – participants first choose a

dative construction in a forced-choice task and are then asked to rate the acceptability of their choice.

The present thesis on the alternation between the Estonian adessive case and the adposition *peal* 'on' makes use of a written rating task. The exact experimental design and some of the practical issues related to it are discussed in section 3.2. In addition, reference is made to other studies conducted by Klavan et al. (2011) that used a forced-choice task and a production task to study the alternation between the Estonian adessive case and the adposition *peal* 'on'. Although I am considering the production task as an experimental study, it should be borne in mind that the kind of data collected with this technique should be considered more similar to corpus data than data collected with a prototypical experiment. Another aspect to bear in mind is that all of the experimental designs discussed are off-line tasks.

1.3.5 Methodological pluralism

Every author who discusses the issue of evidence, data and methodology in linguistics reaches the same logical conclusion – the importance of methodological pluralism, i.e. using a variety of methods to study a specific linguistic phenomenon. No one source of evidence alone can give a final answer (Tuggy 1999: 361) and the value of new data for interpreting old data is proportional to the differences in the methods used to collect it (Labov 1972: 118). The combination of different methods, therefore, increases the reliability of the results and allows for corroboration (Arppe and Järvikivi 2007: 151).

One extremely important, albeit only recent, trend in linguistics is to compare corpus data to experimental data. There is a clear need to corroborate observational data by experimental studies (Dąbrowska 2008: 393), since the advantages and disadvantages of corpora and experiments are complementary. Maybe one of the biggest differences between a prototypically corpus-linguistic study and an experimental design is that while experimental techniques allow for the validation of explicit hypotheses, corpora that give access to natural and spontaneous language tend to be studied only in an exploratory fashion with no rigorous hypothesis testing. Arppe and Järvikivi (2007: 133) in their study on lexical synonymy compare corpus data with a forced choice task and acceptability judgements. Other researchers, e.g. Gries (2002), Featherston (2005a), Bresnan et al. (2007) and Divjak (2008, in progress), compare corpus data only with one set of experimental settings – that of acceptability judgements.

Gilquin and Gries (2009: 17) report in their overview article that the studies that used both corpus and experimental data fall into two types – those that provide converging evidence and those that diverge. Although the importance of being able to account for any differences between corpus and experimental data is fully acknowledged, the relation between the different types of data remains unclear and the comparability of findings is not self-evident. As Arppe and Järvikiv very rightly point out:

> "Although there are obvious benefits in using and combining several sources of evidence, reconciling the different findings with each other presents new challenges, as every method has its own origins and characteristics which all need to be taken into account appropriately. Therefore, this multimethodological development sets new requirements on overall research design and the subsequent argumentation." (Arppe and Järvikivi 2007: 132)

The important question here is to be able to recognise what kind of information different sources of data provide for the linguist to interpret. The one-to-one relationship between different data types should not be taken for granted. Gries (2002: 17) stresses that there are only very few studies that explicitly compare the ways different methodologies lead to different data. Within the ten years that follow Gries's counsel of despair the situation has not improved too much. Among the few studies that do make explicit comparisons, are for example, Kempen and Harbusch (2005), Featherston (2005a), Arppe and Järvikivi (2007), studies referred to in Gilquin and Gries (2009), Divjak (2008, in progress), Bermel and Knittl (in press). Arppe and Järvikivi (2007: 148) conclude that the corpus-based differences do not reflect a real preferential difference, i.e. acceptability. While the forced choice test they employed in their study on two synonymous Finnish verbs fully reflects the results observed in a corpus, the acceptability rating test, although supporting the results of the corpus-based observation and the forced-choice test, also provided a major modification to the original hypotheses. This result mirrors the conclusion of other researchers working on corpus-based frequencies and acceptability ratings that acceptability and frequency do not necessarily correlate with each other (e.g. Divjak 2008, Schmid 2010).

In addition, Arppe and Järvikivi (2007: 150–151) are of the opinion that the two experiments clearly tap into different things: the forced-choice task reflects actual usage situations and is about linguistic performance in production, but the acceptability rating task reflects what is considered possible or appropriate, i.e. it is about introspection. This leads to grouping together forced choice tasks and corpus data in contrast to acceptability/ grammaticality judgements. This is one of the reasons why the present study opted for a rating experiment to complement the corpus studies and not the forced-choice task or the production task. It is assumed that the latter two experimental designs produce similar results as a corpus study, but since rating involves introspection and judgements about what is considered possible and since the data thus acquired differs from corpus data, converging and/or diverging evidence between these two sets of data would be especially beneficial.

It is not necessary for every linguist to use the same set of methods. We do not all need to become psycholinguists or corpus-linguists. What is important, however, is co-operation. The only road to resolving theoretical aspects about controversial issues in linguistics is comparing and interpreting data that is truly comparable, i.e. not based on personal preferences as pertains to intuition and introspection. On the other hand, when we are looking at a specific phenomenon, it is the responsibility of the linguistic community to study this pheno-

menon by gathering data from a variety of distinct sources. The aim of the present thesis is to provide data from at least two distinct sources – written corpus data and rating tasks. It is clear that there are important aspects that are not considered here (e.g. spoken observations, psycholinguistic experimentation as pertains to processing differences between analytic and synthetic forms, language acquisition, language change, etc.). It is hoped that future research will undertake the study of the alternation between locative cases and adpositions along these lines to ensure that variety of methods are used to "converge on right answers to hard questions" (Labov 1972: 119).

2. EXPRESSING SPACE IN ESTONIAN: LOCATIVE CASES AND ADPOSITIONS

The previous chapter outlined the theoretical assumptions underlying the study and discussed some of the aspects of linguistics as an empirical science. The most important conclusion is that the combination of different data sources is necessary for a linguistic analysis. Before moving on to the design of the studies and the analysis itself, it is necessary to describe in detail the linguistic phenomenon under study. Therefore, this chapter gives an introduction to the Estonian adessive case and the adposition *peal* 'on'. It starts with a general discussion of how space is expressed in Estonian and looks more specifically at the alternation between the two locative constructions studied. The second part of the chapter is a non-exhaustive discussion of several studies on various other alternation phenomena, mainly the alternation between the *s*- and *of*-genitive, particle placement and dative alternation in English, and their basic conclusions. These studies serve as the background and an inspiration for the present corpus and experimental studies. The chapter ends with a list of both the semantic and morphosyntactic variables included in the dissertation and highlights some of the reasons why other variables are not included. The last section spells out the specific objectives and general predictions of the corpus and experimental studies, serving thus as an introduction to Chapter 3 which describes the selection and the implementation of the specific methodology.

The present dissertation adopts Langacker's (1987, 2008) terminology to refer to the two most fundamental notions in discussing language expressions – Trajector and Landmark. The Estonian adessive case construction and the adpositional construction with *peal* 'on' are both relational expressions, i.e. they profile relationships between their participants (Langacker 2008: 112–127). The most prominent participant is called the Trajector and the second participant is the Landmark (Langacker (2008: 70). **Trajector** is the entity whose location or motion is of relevance; **Landmark** is the reference entity in relation to which the location or the motion of the Trajector is specified. Trajector may be static or dynamic, a person or an object, or even a whole event. The following illustrative examples are taken from Zlatev (2007: 327; the Trajector is underlined):

(3) a. *She is at school.* = static
b. *She went to school.* = dynamic
c. *The book is on the table.* = object
d. *She is playing in her room.* = whole event

The concepts of Trajector and Landmark are closely related to Talmy's (1983, 2000) notions of Figure and Ground. Talmy (1983: 230–231, 2000: 315–316) identifies certain object properties that favour the Figure or Ground construal. Figure is defined as having unknown spatial (or temporal) properties to be determined, and Ground acts as a reference entity, having known properties that

can characterize the Figure's unknown properties (Talmy 2000: 315). In addition, it is characteristic of a Figure to be more movable, smaller, geometrically simpler, more recently on the scene, of greater relevance, less immediately perceivable, more salient and more dependent than a Ground (Talmy 2000: 315–316). The concepts of Figure-Ground and Trajector-Landmark largely pertain to the same phenomena. Although in spatial language studies, the terminology of Figure-Ground is employed more frequently, the present dissertation makes use of Langacker's (1987, 2008) terminology. I agree with Langacker (1987: 231–237) who presents arguments for treating the Trajector-Landmark asymmetry as a special case of Figure-Ground alignment. The Figure-Ground alignment is taken to refer to a more general mechanism in human cognition, while Trajector-Landmark is taken to apply more specifically to linguistic expressions. The point of departure in the present study is the linguistic expression (i.e. the two constructions) which leads to the underlying conceptualisation and not vice versa; the latter being perhaps more frequent in studies placing more emphasis on spatial cognition.

The third relevant notion in discussing spatial expressions in cognitive linguistic terms is **region**. It has been suggested that languages do not relate the Trajector and Landmark in a spatial expression directly, but through a "region" (Landau and Jackendoff 1993, Svorou 1994, Zlatev 2007). Although the concept of region has been mentioned in several theories of spatial relations (see also Langacker 1987: 198), it was Svorou (1994) who fully developed this notion and gave it conceptual priority. In essence, the term region refers to "an area adjacent to a [Landmark] (or part of it) in which a specific spatial description is valid" (Svorou 1994: 13). The concept of region is claimed to be especially relevant within a theory of spatial relations which assumes a relativistic idea of space, i.e. space is understood by the relations that exist between objects, and where knowledge about the size, mobility, and interactional and functional attributes of entities play a role (Svorou 1994: 15). Zlatev also claims that "most, if not all, of the regions that are relevant for spatial semantics correspond to various types of "image schemas" such as containment and support" (2007: 330). Image schemas are imaginative structures of understanding which provide a conceptualization derived from perception and bodily experience. According to Johnson:

> "An image schema is a recurring, dynamic pattern of our perceptual interactions and motor programs that gives coherence and structure to our experience. One of the central arguments of this book is that experientially based, imaginative structures of this image-schematic sort are integral to meaning and rationality." (Johnson 1987: xiv)

It is important to note that image schemas are not specific images but are schematic. They present schematic patterns arising from our physical experience of being and acting in the world, e.g. moving our bodies, exerting force, etc. Johnson showed that image schemas "are pervasive, well-defined, and full

of sufficient internal structure to constrain our understanding and reasoning" (1987: 126). They are somewhat more basic than, for example, the higher level conceptual structure of metaphor. The present thesis studies some of the ways how the support-relation is expressed in Estonian. As a working definition, this term denotes a spatial scene where the Trajector is located on top of a Landmark.

2.1 Description of the linguistic phenomenon

Estonian has a variety of means available for expressing spatial relations. The two most common ways to talk about space in Estonian are the locative cases and the numerous adpositions. Estonian reference grammars list the following nine cases out of the total 14 as locative: illative, inessive, elative, allative, adessive, ablative, translative, terminative, and essive[19]. The entire case paradigm for the Estonian language is given in Table 2 with the locative cases in bold.

Table 2. Case formation in Estonian

Case	Singular	Plural
nominative	hoov 'yard'	hoovid 'yards'
genitive	hoovi 'of yard'	hoovide 'of yards'
partitive	hoovi 'yard'	hoovisid/hoove 'yards'
illative	hoovisse/hoovi 'into yard'	hoovidesse 'into yards'
inessive	hoovis 'in yard'	hoovides 'in yards'
elative	hoovist 'out of yard'	hoovidest 'out of yards'
allative	hoovile 'onto yard'	hoovidele 'onto yards'
adessive	hoovil 'on yard'	hoovidel 'on yards'
ablative	hoovilt 'from yard'	hoovidelt 'from yards'
translative	hooviks 'into yard'	hoovideks 'into yards'
terminative	hoovini 'until yard'	hoovideni 'until yards'
essive	hoovina 'as yard'	hoovidena 'as yards'
abessive	hoovita 'without yard'	hoovideta 'without yards'
comitative	hooviga 'with yard'	hoovidega 'with yards'

The locative cases belong to the group of **concrete** cases which are opposed to **abstract** cases in Estonian reference grammars (Erelt et al. 1993: §21, Erelt et al. 2007: 240). Nominative, genitive, and partitive belong to the set of abstract cases while the rest of the 11 cases are considered concrete. The authors of the Estonian reference grammars point out that what differentiates abstract cases from the concrete cases is that while the meaning of the abstract cases does not

[19] Cases like the essive and translative are listed here as locative cases for reasons of language history. It is claimed that the earlier forms of the essive and translative case endings were originally the same as the case ending for the locative case (Rätsep 1979: 63–74).

depend on the lexical meaning of the noun inflected for these cases, the primary meaning of the concrete cases can be defined only together with the lexical meaning of the noun to be inflected. The treatment of case in Estonian reference grammars is broadly similar to Chomsky's treatment of case (Chomsky 1982, 1986, 1988). Case is seen as a predominantly syntactic phenomenon – both case and its governing elements are considered "semantically empty" and treated as purely grammatical phenomena. Case endings are meaningless items which are required in certain structural positions or governed by particular lexical items. Although the aim of the present thesis is not to present a bullet-proof argument for considering the grammatical category of "case" as meaningful in its own right, it is the firm belief within cognitive linguistics that both abstract and concrete cases are meaningful in their own right and should be treated as such in reference grammars as well. A reader sympathetic to this line of argumentation is referred to the cognitive-functional studies on (grammatical) cases (e.g. Dąbrowska 1997; Jaakola 2004; Janda 1993, 2002, 2006) which convincingly demonstrate that even grammatical cases like the Czech and Polish dative, the Czech and Russian instrumental, and the Finnish genitive, traditionally considered as abstract and void of meaning, have meanings in their own right. It is therefore assumed in this dissertation that the more abstract functions of the adessive case listed below in section 2.1.1 are in fact systematically related to its spatial functions. This kind of semantic systematicity between the abstract and spatial functions is presumed for the other locative cases as well.

Contrary to the comprehensive list of locative cases, providing such a list for Estonian adpositions is difficult. For example, Grünthal (2003: 56) demonstrates that in different grammatical descriptions and lexical overviews the number of adpositions in Estonian varies greatly and depends on the way the category of an adposition is defined; a problem that exists in many other languages as well. In general terms, adpositions can be defined as "free morphological forms that appear in languages primarily in a construction with noun phrases, either preposed (prepositions) or postposed (postpositions) to indicate case and case-like functions such as space, time, causality, or instrument" (Svorou 2007: 726). Although the existence of a class of adpositions is in general accepted, problems arise as soon as it is attempted to define the category, establish a list of its members, and mark its boundaries (Dryer 2007). Different linguists adopt different criteria and the descriptions can thus depart from each other significantly. It has also been pointed out that it would be wrong to accept "adposition" as a well-defined universal category, since functionally equivalent terms to adposition like *co-verb, verbid, relational nouns* have been proposed for typologically different languages that do not quite fulfil all the requirements for an adposition, but do participate in constructions where they play the role of an adposition (Svorou 2007: 727).

It is interesting to note that while the English descriptive grammars (e.g. Quirk et al. 1985) give at least approximate numbers for the category of adpositions (cf. Landau and Jackendoff 1993 who estimate the figure to be about 80 to100), no such list can be found in the Estonian reference grammars

(Erelt et al. 1993, 1995). However, other sources give the following numbers for Estonian: Stoebke (cited in Grünthal 2003: 56) gives 89 adposition stems for Estonian, Tauli (cited in Grünthal 2003: 56) lists 140 postpositions, Palmeos (1985) around 116 adpositions, and the Written Dictionary of Estonian (cited in Grünthal 2003: 56) gives a somewhat larger number of 185 adpositions in total, 135 are exclusively postpositions and 29 exclusively prepositions, and 19 are bipositionals[20].

The main reason why no comprehensive list of (Estonian) adpositions can be provided has to do with problems determining word classes. Although it is commonly claimed that spatial meaning is cross-linguistically expressed by members of the closed classes of words (Svorou 1994, Talmy 2000), the status of Estonian adpositions, as in many other languages, is not as black and white – they can be considered a class of words neither completely open nor completely closed, but having characteristics of both types of word classes and thus falling somewhere in the middle on the scale from completely closed word classes to completely open word classes. According to Schachter and Shopen (2007: 3) open classes are those whose membership is in principle unlimited and closed classes those that contain a fixed and usually small number of member words. As we saw in the previous paragraph, the number of Estonian adposition is neither fixed nor is it particularly small (Grünthal 2003: 56). This seems to weigh the scales towards Estonian adpositions being an open class of words. However, since adpositions in Estonian, similarly to new members of other closed classes, are a product not of derivation from other elements, but rather of evolution or grammaticalization processes (Svorou 1994: 31, Lehmann 2002: 119), the scales tilt in the other direction. As is appealing to any linguist familiar with prototype theory (Rosch 1973, Rosch et al. 1975, 1976), the distinction between the open and closed word classes is gradual (Lehmann 2002: 119). Whether a word already belongs to the closed class or is still in the open class depends on degree of grammaticality.

2.1.1 The Estonian adessive case

The Estonian adessive case is one of the nine cases referred to as locative cases in Estonian. It belongs to the set of **external locative cases** that express spatial relations of an open surface and they form a three-part series – allative, adessive, ablative – expressing direction, location and source respectively (Erelt et al. 2007: 240; see Table 3). The adessive takes the case ending of *–l* as in *hoovi-l* ‘on yard’. The other members of the set of external locative cases take the endings of *–le* in the allative, as in *hoovi-le* ‘onto yard’, and *–lt* in the ablative, as in *hoovi-lt* ‘off yard’. The second set of locative cases is referred to as the **interior locative cases** and comprises the illative (*-sse* as in *hoovi-sse* ‘into yard’), the inessive (*-s* as in *hoovi-s* ‘in yard’), and the elative (*-st* as in

[20] The term *bipositional* refers to syntactically ambiguous adpositions which may occur either as prepositions or postpositions (Grünthal 2003: 46).

hoovi-st ‘out of/ from yard’) as exemplified in Table 3. The lative members of the locative cases, i.e. the allative and illative express direction, the locative members (adessive and inessive) express location, and the separative members (ablative and elative) express the source. This section focuses only on the adessive case.

Table 3. Formation of Estonian locative cases as exemplified by the noun *hoov* ‘yard’

	LATIVE (direction)	LOCATIVE (location)	SEPARATIVE (source)
Interior	illative *hoovi-**sse*** ‘into yard’	inessive *hoovi-**s*** ‘in yard’	elative *hoovi-**st*** ‘out of/ from yard’
Exterior	allative *hoovi-**le*** ‘onto yard’	adessive *hoovi-**l*** ‘on yard’	ablative *hoovi-**lt*** ‘off yard’

The Estonian adessive case normally takes the role of an adverbial (as in *laual* ‘on the table’ in example 4) or adverbial modifier (as *vaas laual* ‘vase on the table’ in example 5) (Erelt et al. 1995: 58).

(4) *Vaas* *on* ***laual***.
vase.SG.NOM be-PRS.3SG table.SG.ADE
‘The vase is **on the table**.’

(5) ***Vaas*** ***laual*** *on* *ilus*.
vase.SG.NOM table.SG.ADE be-PRS.3SG pretty.SG.NOM
‘**The vase on the table** is pretty.’

According to Estonian reference grammar (Erelt et al. 2007: 250) the function of the Estonian adessive case is to express location (e.g. *laual* ‘on the table’ in example 6a), temporal relations (e.g. *neljapäeval* ‘on Thursday’ in example 6b), states (e.g. *naerul* ‘with a smiling’ in example 6c), possessors in possessive clauses (e.g. *Maril* ‘Mari’ in example 6d), agents with finite verb forms (e.g. *mul* ‘I’ in example 6e), instruments (e.g. *klaveril* ‘on a piano’ in example 6f), and manner (e.g. *kõrvul* ‘with ears’ in example 6g).

(6) Functions of the Estonian adessive case (Erelt et al. 2007: 250):

a. Location: *Vaas* *on* ***laual.***
vase.SG.NOM be-PRS.3SG table.SG.ADE
‘The vase is **on the table**.’

b. Time: *Nad* *sõidavad* ***neljapäeval*** *maale.*
they.NOM drive-PRS.3PL Thursday.SG.ADE country.SG.ALL
‘They are driving to the country **on Thursday**.’

c. State: *Jüri* *vaatas* *meid* ***naerul*** *näoga.*
Jüri.NOM look-PST.3SG us laugh.SG.ADE face.SG.COM
'Jüri looked at us **with a laughing** face.'

d. Possessor: ***Maril*** *on* *kaks* *last.*
Mari.ADE be-PRS.3PL two child.SG.PRT
'**Mari** has two children.' (lit. '**On Mari** are two children.')

e. Agent with finite verb forms:
See *asi* *ununes* ***mul*** *kiiresti.*
this.SG.NOM thing.SG.NOM forget-PRS.3SG me.SG.ADE quickly
'**I** quickly forgot about that thing.'

f. Instrument: *Mari* *mängib* ***klaveril*** *mõnd lugu.*
Mari.NOM play-PRS.3SG piano.SG.ADE some tune.SG.PART
'Mari is playing some tunes **on the piano**.'

g. Manner: *Mari* *kuulas* *kikkis* ***kõrvul.***
Mari.NOM listen-PST.3SG pricked.up ear.PL.ADE
'Mari listened **with her ears** pricked up.'

As can be seen from the examples given in 6, the Estonian adessive case fulfils many functions besides location and many of them are relatively abstract. As the corpus results presented in Chapter 4 demonstrate, it is more frequent for the Estonian adessive to express the possessor or agent (examples 6d and 6e) than, for example, location (example 6a). This has led some linguists working on Estonian to object to referring to the adessive as a locative case (Matsumura 1994, 1996, 1997). However, according to the localist theory the concrete uses of a case are more primary than the more abstract uses (Andersson 1971, 2006: 95–96; Lyons 1977: 718–724). It can be argued that expressing location is still one of the most important functions of the Estonian adessive case although the raw frequencies of a corpus analysis show different results. For the present study, only the adessive functions of expressing location (example 6a) and instrument (example 6f) are relevant.

Although the focus of the dissertation is only on the spatial uses of the adessive case, since this is where the adposition *peal* 'on' provides a valid alternative, the other functions of the adessive cannot be left out of the picture. Polysemy of a grammatical construction influences the synonymous relationships this construction can enter into with other grammatical constructions. The adessive use in the possessive construction for marking the possessor is considered to be one of the most important functions that interacts with the adessive use in the locative construction. It is assumed that one of the reasons why the *peal*-construction is used instead of the adessive construction in expressing a support relation is to avoid ambiguity between the possessive and locative function of the adessive case. However, a detailed study on how polysemy exactly interacts with synonymy between the two constructions is left for future research.

Estonian external locative cases have been studied from the perspective of Cognitive Grammar by Vainik (1995). Vainik (1995) adopts a Langackerian-style analysis of the semantics of the locative cases. Unlike more traditional work on case inflections (see, for example, the references in Janda 1993 and Dąbrowska 1997) the meaningfulness of case-endings is taken as self-evident in such cognitively oriented work on cases. Vainik's (1995) work is similar to the cognitive linguistic work on case inflections conducted by Janda (1993) and Dąbrowska (1997) – the aim is to account for all of the different uses of the external locative cases via concepts of prototypes and radial semantic networks. Although the value of Vainik's (1995) very detailed and qualitative study is unquestionable, no direct comparisons can be made with the studies reported here. Vainik does not directly address the issue of the alternation between locative cases and the corresponding adpositions, nor does it present much quantitative empirical data to support the theoretical claims. Validating the claims made in Vainik (1995) falls outside the scope of this dissertation, since the focus is on synonymy and not on polysemy.

A more quantitatively oriented work on the Estonian adessive is Laaksonen (2002). She compares the use of the Estonian adessive to the Finnish adessive using corpus-linguistic methods. The study limits itself to providing frequency counts and no statistical measures are provided that assess the validity of the results. The focus in Laaksonen (2002) is again on looking at the adessive construction in general and the locative function only forms a minor part in this study. Laaksonen (2002) reports that the most frequent use of the adessive in her corpus of 1009 occurrences of the Estonian adessive in fiction texts involves expressing the possessor or an agent with the finite verb form (examples 6d and 6e above), followed by the adessive used in the locative construction. The third most frequent use is that of temporal relations (example 6b above). The corpus data (3680 adessive occurrences in fiction texts) studied in this dissertation elicits a slightly different ranking. The most frequent function of the adessive is expressing time, followed closely by expressing the possessor or another logical subject. The spatial use of the adessive is approximately three times less frequent than the two abstract functions (see section 3.1.1 for details). Naturally, the two corpus studies differ in that they look at two different corpora and use different criteria for distinguishing between the different functions of the adessive. The more abstract functions of the Estonian adessive have been treated by Matsumura (1994, 1996, 1997), Lindström and Tragel (2007, 2010). No direct comparisons are made at this point between the results of the studies that focus on the more abstract uses of the adessive and the studies reported here for reasons stated above.

2.1.2 The Estonian adposition peal 'on'

In addition to the locative cases, location and change of location in Estonian can be expressed with adpositions and adverbs. One of the primary ways in which languages differ from one another is in the relative ordering of subject (S), verb (V), and object (O) (Dryer 2007: 61). The Estonian language, like other Finnic languages, has presumably changed from a historical SOV to SVO and is predominantly postpositional (Grünthal 2003: 45). In fact, Estonian adpositions are interesting in this respect since they have a typologically "double character" (Grünthal 2003: 45), i.e. there are both prepositions and postpositions in Estonian. According to both Grünthal (2003:45) and Dryer (2005) mixed adpositional systems are exceptional in the world's languages. Although there are both prepositions as well as postpositions in Estonian, Grünthal (2003:45) has pointed out that the number of prepositions is rather small and does not exceed 20-25% of all adpositions.

The Estonian adpositional phrase consists of an NP and a pre- or postposition (Erelt et al. 1993: 137). Estonian adpositional phrases, especially prepositional phrases are exocentric, because neither of the two constituents can be omitted (Erelt et al. 1993: 137, Grünthal 2003: 47). According to Grünthal (2003: 47) postpositional phrases are also exocentric, but they are syntactically more flexible than prepositions. Grünthal (2003: 63) emphasises that although prepositions and postpositions belong to the same category and the same items may in certain cases even be used both as prepositions and postpositions, their syntactic location and relation with respect to the noun differ in many ways. Table 4 presents the main morphosyntactic characteristics of Estonian prepositional and postpositional phrases. Tauli (1966: 44) has also proposed that the meaning of the prepositions is often more abstract and that of the postposition more concrete.

Table 4. Morphosyntactic characteristics of Estonian prepositional and postpositional phrases (Grünthal 2003: 65)

Prepositions	Postpositions
• predominantly partitive-governing	• predominantly genitive-governing
• low degree of inflection	• higher degree of inflection
• occasional case inflection	• case inflection to some extent
• no possessive suffixes	
• prevailing semantic roles: PATH, CIRCUMSPATIAL	• prevailing semantic roles: SPATIAL
• additional NP determiners such as pronouns and attributes may be located between the two components of PrepP	• no free word may be added between the noun and the postposition

The Estonian adposition *peal* 'on' is a postposition and requires the head noun to be in the genitive case. In addition, it is important to note that no free word may be added between the noun and the postposition *peal* 'on'.

In defining the category of adpositions in Estonian, it is commonly stated that adpositions are uninflected words which belong together with a nominal and express different relations with that nominal (Palmeos 1985: 3, Erelt et al. 1995: 33). A distinctive morphological characteristic of Estonian adpositions is that like adverbs and particles they constitute three-member sets that are semantically and grammatically divided into the lative, locative, and separative form. The lative member expresses direction and takes either an illative or allative case ending; the locative member expresses location and takes either an inessive or adessive case ending; the separative member expresses direction and takes an elative or ablative case ending. As can be seen from Table 5, the adposition *peal* 'on' takes external locative case endings: *pea**le** – pea**l** – pea**lt*** and the adposition *sees* 'in' interior locative case endings: *si-**sse**, see-**s**, see-**st***.

Table 5. Three-member sets of Estonian adpositions *sees* 'in' and *peal* 'on'

	LATIVE (direction)	LOCATIVE (location)	SEPARATIVE (source)
Interior	illative *si-**sse*** 'into'	inessive *see-**s*** 'in'	elative *see-**st*** 'out of'
Exterior	allative *pea-**le*** 'onto'	adessive *pea-**l*** 'on'	ablative *pea-**lt*** 'off'

Grünthal (2003: 74) notes that 83% of unambiguously genitive-governing Standard Estonian postpositions carry a productive local case ending. He states that "considering the fact that the interior local cases (illative, inessive, elative) denote more concrete spatial relations than the exterior ones, it is somewhat surprising that Estonian postpositions most commonly display the adessive" (Grünthal 2003: 74). A very interesting research topic would be to study whether and how the case endings influence the meaning of Estonian post-positions, i.e. does it somehow reflect in their semantics which case endings, either interior or exterior, they have affixed during the course of grammati-calization. Unfortunately, this issue cannot be further discussed in the present thesis, but see Lestrade (2010), Ojutkangas and Huumo (2010) for studies on a similar topic in Finnish.

Another prominent characteristic of (Estonian) adpositions is that they can belong to various grammatical categories – the same linguistic item, e.g. *peal* 'on', can be realized as an adposition or as an adverb. This leads to problems in determining the word class of spatial grammatical words, as pointed out by, e.g. Dryer (2007) and Veismann (2008, 2009). The present study focuses on the use of *peal* as an adposition; the use of this linguistic item as an adverb and particle has been discarded. The question of how the membership of a lexical unit in a

certain word class influences its meaning is of course interesting, but this issue is outside the scope of the present thesis. Similarly to Veismann (2008, 2009), I agree with O'Dowd (1998) who has shown that the realization of these three word classes (adpositions, adverbs, and particles) depends on discourse-functional factors.

At the clause level, the Estonian adpositional phrase has two basic functions, that of an adverbial (as *laua peal* 'on the table' in example 7) and adverbial modifier (as *vaas laua peal* 'the vase on the table' in example 8). As is seen above (examples 4 and 5), the adessive case fulfils the same two basic syntactic functions.

(7) *Vaas* *on* ***laua*** ***peal***.
vase.SG.NOM be-PRS.3SG table.SG.GEN on
'The vase is **on the table**.'

(8) ***Vaas*** ***laua*** ***peal*** *on* *ilus*.
vase.SG.NOM table.SG.GEN on be-PRS.3SG pretty.SG.NOM
'**The vase on the table** is pretty.'

Similarly to locative cases, Estonian adpositions are also polysemous[21]. The Dictionary of Written Estonian (Langemets et al. 2009: 130–131) lists as many as 11 meanings for the adposition *peal* 'on'; relevant for the present study are the functions given in examples 9a–9c. Similarly to the functions of the adessive, it is assumed in the dissertation that the more abstract functions of the adposition *peal* 'on' are systematically related to its spatial functions.

(9) Functions of the Estonian adposition *peal* 'on' (Langemets et al. 2009: 130–131):

a. Location: *Leib* *on* ***laua*** ***peal***.
bread.SG.NOM be-PRS.3SG table.SG.GEN on
'Bread is **on the table**.'

b. Place: ***Turu*** ***peal*** *oli* *suur* *sagimine.*
market.SG.GEN on be-PST.3SG big.SG.NOM commotion.SG.NOM
'There was a big commotion **on the market**.'

c. Instrument: *Mängi* ***klaveri*** ***peal*** *ette!*
play-IMP.2SG piano.SG.GEN on ahead
'Perform it **on the piano**!'

When comparing the list of the functions in example 9 to the list of the functions provided for the adessive case in the previous section in example 6, it can be seen that the two constructions are valid alternatives for specifically the three functions in 9a–9c. However, as with the adessive case, the adposition

[21] The polysemy of Estonian adpositions from a cognitive linguistics perspective has been studied, for example, by Veismann (2006, 2008, 2009).

peal ‘on’ also expresses relations for which the adessive case is not a valid alternative. A polysemous account of the adpositional construction in general is necessary, but falls outside the scope of the present dissertation. As mentioned earlier, the focus in the dissertation is on the synonymy between the two constructions. It is believed that the polysemy of the adposition *peal* ‘on’ is less influential in the alternation between the two constructions than the polysemy of the adessive case. This claim needs empirical validation and future research hopefully takes polysemy of the two constructions into account when studying the alternation between them.

2.1.3 Alternation between the Estonian adessive and the adposition peal ‘on’

The synthetic adessive case construction and the analytic adpositional construction with *peal* ‘on’ are used as alternatives when expressing location and instrument. According to Palmeos (1985: 15), the analytic construction – the genitive together with the adposition *peal* ‘on’ – expresses the same meaning as the synthetic adessive. At the same time, it has been claimed in Estonian reference grammars that the meaning of adpositions is more concrete and specific than that of the cases (Erelt et al. 2007: 191). In comparison with adpositions, the meaning of cases is said to be much more abstract and the usage range much broader (Erelt et al. 1995: 33–34, Erelt et al. 2007: 191). This is in line with the general claims made in literature concerning the differences between adpositions and case affixes (Comrie 1986, Hagège 2010, Lestrade 2010). In spite of such broad claims, no detailed and systematic quantitative study exists on the topic – the alternation between Estonian synthetic locative cases and the corresponding analytic adpositional constructions.

Despite the difference between the two constructions, there are still instances where both the adessive case and the adposition *peal* ‘on’ are seen as semantic alternatives in Estonian. The two constructions can be said to be synonymous because they render the same content (cf. examples in 6a and 9a) – both express a spatial location between two entities, where one is placed on top of the other, i.e. the support-relation. In Langacker’s (1987) terminology, they profile the same relationship. As we will see later in Chapter 4, distinctive collexeme analysis shows that in the 5-million-word fiction sub-corpus of the Balanced Corpus of Estonian, there are as many as 182 different lexemes used with both the adessive case and the adposition *peal* ‘on’. Further converging evidence that the two constructions are used as alternatives comes from various open and semi-open production tasks. Klavan et al. (2011) report the results of a forced choice task and a production task the aim of which was to determine which semantic variables play a role in the use of the adessive and adposition *peal* ‘on’. In addition to demonstrating that there are semantic differences between the two constructions, the two tasks also yielded results where there was no significant difference between the two locative constructions. There is evidence for claiming that the two constructions are near-synonyms. On the very general

level they both express a support-relation, but it is hypothesised that the variation between the synthetic adessive construction and analytic *peal*-construction is not free. Instead, the variation is assumed to be conditioned by a set of semantic and morphosyntactic variables discussed in section 2.3 below.

2.1.4 Alternation between other locative cases and adpositions in Estonian

In addition to the alternation between the adessive case and the adposition *peal* 'on', the Estonian language also exhibits parallel uses between the other locative cases and the corresponding adpositions, as in examples 10a–10b and 11a–11c below:

(10) a. *Paneb* *raamatu* { ***lauale*** / ***laua*** ***peale***. }
put-PRS.3SG book.SG.GEN table.SG.ALL table.SG.GEN onto
'He/She puts the book **on(to) the table**.'

b. *Võtab* *raamatu* { ***laualt*** / ***laua*** ***pealt***. }
take-PRS.3SG book.SG.GEN table.SG.ABL table.SG.GEN from on
'He/She takes the book **from the table**.'

(11) a. *Paneb* *raamatu* { ***kasti*** / ***kasti*** ***sisse***. }
put-PRS.3SG book.SG.GEN box.SG.ILL box.SG.GEN into
'He/She puts the book **in(to) the box**.'

b. *Raamat* *on* { ***kastis*** / ***kasti*** ***sees***. }
book.SG.NOM be-PRS.3SG box.SG.INE box.SG.GEN in
'The book is **in the box**.'

c. *Võtab* *raamatu* { ***kastist*** / ***kasti*** ***seest***. }
take-PRS.3SG book.SG.GEN box.SG.ESS box.SG.GEN from in
'He/She takes the book **from the box**.'

Similarly to the alternation between the adessive and *peal*, the alternations exhibited in examples 10a–10b and 11a–11c have so far not been analyzed in sufficient quantitative detail, save for the alternation between the allative and the adposition *peale* 'onto' (example 10a). The latter was studied by Kesküla (2011) who employed both univariate (chi-squared test) and multivariate (logistic regression modelling) corpus analysis to study the effect of various semantic and morphosyntactic variables on the 300 allative and 300 *peale* 'onto' occurrences. Kesküla (2011) reports that the adpositional *peale*-construction tends to be used when the locative phrase holds a middle position within the clause, the Trajector is a thing or an activity, the Landmark is animate and pronominal, and when there is meaning transfer; the allative construction tends to be used when the Trajector is plural. The results of Kesküla's (2011) study are compared to the results obtained in the corpus

studies reported in the dissertation and Chapter 4 provides a more detailed discussion on this comparison.

Although this kind of alternation of Estonian synthetic and analytic locative constructions is a typologically intriguing language phenomenon, there are to date no detailed studies on this topic, excluding very few small-scale studies. One such study was conducted by Rannat (1991) who analysed the Estonian locative cases (both interior and exterior) and all of the possible adpositions that can be used as alternatives to them. Among Rannat's (1991) dataset the synonymous use of the Estonian adessive and the adposition *peal* 'on' constitutes just a very small part. Rannat (1991) concludes, based largely on her own intuition, that the preference of either the synthetic or analytic form does not depend on the syntactic composition of the clause (e.g. the transitivity of the predicate verb) and that it may depend instead, for example, on the clause stress (Rannat 1991: 15). In addition, Rannat (1991: 52–54) points out that idiolect may play a role. Although I do agree that these variables may influence the use of the adessive and the adposition *peal* 'on', the (morpho)syntactic variables also merit a more detailed analysis.

In addition, in other Finno-Ugric languages, Bartens (1978) and Ojutkangas (2008) have looked at the alternation between the interior locative cases and the corresponding adpositions in the Saami and Finnish languages respectively. The central claim of Bartens (1978) is that the analytic adpositional construction places more stress on the location than the synthetic case construction. In addition, Bartens (1978) specifies that in the Saami languages the adpositional constructions are used together with smaller, manipulable things as Landmarks as well as with vehicles. Ojutkangas (2008) reports the results of a corpus study based on Finnish dialects and her central claim is that the interior locative cases express conventional spatial relations between the Trajector and Landmark, while the corresponding adpositional constructions are used when this relation is somewhat unconventional. The comparison of the results reported here with the studies conducted by Bartens (1978) and Ojutkangas (2008) is given in Chapter 6.

2.2 Studies on other alternation phenomena and the methodologies used

This section takes a look at some of the relevant previous studies on various alternation phenomena and the methodologies employed to study them. Focus will be on the type of methodology used (corpus studies vs. experiments), variables included and the results obtained. The majority of the studies reviewed focus on comparing two synonymous constructions or lexemes; Arppe (2008: 17) suspects that the pairwise comparisons of synonyms is methodologically the easiest to pursue. However, as Arppe (2008) himself shows in his study and as Divjak (2010) argues, synonymy as a phenomenon is by no means restricted to pairs. For example, in addition to looking simply at the alternation

between the adessive and *peal* 'on', one could also include adpositions like *otsas* 'on top of' and *küljes* 'attached to' or other locative cases such as the inessive in the analysis. It was, however, decided to limit the present study only to the adessive and *peal* 'on', because the alternation between these two constructions is more frequent.

In addition to dividing the relevant previous studies into two groups based on whether they cover grammar or lexicon, a different division is possible based on the approach taken. As Gries (2003a: 162) and Hinrichs and Szmrecsanyi (2007: 438) point out, the trend in the more traditional approaches to variation is to take a univariate approach by considering only univariate results and disregarding the fact that for language users all variables are present simultaneously. Some recent examples of studies that are concerned with isolating a certain effect and taking a univariate rather than multivariate approach include Rosenbach (2005, 2008) on the English genitive variation and Antilla et al. (2010) on the role of prosody in the English dative alternation. Recent years, however, have seen the boom of truly multivariate studies: Gries (1999, 2001, 2003a, 2003b) on the English particle placement; Bresnan et al. (2007), Bresnan and Ford (2010) on the English dative alternation; Szmrecsanyi (2005, 2009, 2010), Hinrichs and Szmrecsanyi (2007), Szmrecsanyi and Hinrichs (2008), Shih et al. (to appear), Wolk et al. (to appear) on the English genitive alternation; Arppe (2008) on Finnish cognition verbs, Divjak (2006, 2008, 2010), and Divjak and Gries (2006) on Russian near-synonymous verbs, to name just a few. These multivariate studies acknowledge that variation is governed by an intricate interplay of several variables and that univariate analysis techniques are inappropriately reductionist and simplistic. Although Rosenbach (2005: 639) argues that it is not necessary to investigate all the possible variables simultaneously, the crucial point is that researchers should be aware of the presence of other variables, especially if these variables are strongly correlated.

2.2.1 Studies on grammatical synonymy

Pairs of semantically equivalent expressions (referred to as *alternating pairs*) such as the English dative and genitive alternation, active and passive voice, and verb-particle constructions have played and continue to play an important role in the development of both formal and functional theoretical paradigms (Gries and Stefanowitsch 2004b: 97–98). In the generative grammar tradition the relationship between the members of alternating pairs is captured in terms of derivation mechanisms relating both members of a pair to the same underlying structure (e.g. Chomsky 1957, Levin and Rappaport Hovav 1995). Discourse-functional approaches, on the other hand, have focused on the different ways these members package information flow discussing notions like topicality, thematicity, givenness or animacy (e.g. Givón 1993). While none of the previously mentioned approaches consider semantics, the latter plays a fundamental role in the more recent construction-based approaches to language. Approaches like Langacker's (1987) Cognitive Grammar or Goldberg's (1995,

2002, 2006) Construction Grammar do not consider one of the members of the alternating pair as basic and the other as derived, or deriving both from the same underlying source – each member is a construction in its own right. A word may alternate between two constructions if the word's meaning is compatible with the meanings of both constructions.

The fact that alternations still capture the attention of researchers is confirmed by the sheer number of edited books, monographs, journal articles, and conference papers dedicated to the subject. The majority of the studies discussed in this section focus on one specific alternation, but there are a few exceptions like Gries and Stefanowitsch (2004b), Gries (2007b), Szmrecsanyi (2005, 2009, in press a), Wolk et al. (to appear) that take a look at alternation phenomenon in general or factor in more than one alternation at a time. Wolk et al. (to appear) studied the development of two cases of syntactic variation – the genitive and dative – during the Late Modern English Period. In comparison to the majority of variation studies that focus on individual cases of syntactic variation, Wolk et al.'s (to appear) corpus study provides a large-scale quantitative study of two alternations and compares how they develop over time. Such an approach allows the researchers to assess whether the same effects for the same variables are present in both types of alternation or whether there are differences across different alternation phenomena. Wolk et al.'s (to appear) data demonstrates a stable preference over time for placing animate referents first and the short-before-long preference; what changes is the strength of the effects. As Wolk et al. (to appear: 27) themselves stress "[a]dopting a cross-constructional approach to syntactic variation and change may thus point to general changes in grammar which could remain elusive when looking at some specific alternation in isolation". A further important aspect about Wolk et al.'s (to appear) study is that they employ mixed effects modelling, allowing for idiolectal and lemma-specific random effects (see section 3.4 for an overview of mixed-effects modelling).

Another study that looks at more than one alternation simultaneously is Szmrecsanyi (2005) who analyses three well-known alternations in spoken corpora of English: analytic vs. synthetic comparatives, particle placement, and future marker choice. Szmrecsanyi (2005) concludes that factoring in persistence in addition to the more commonly discussed variables like length, complexity, frequency, definitess, and topicality increases the researcher's ability to account for linguistic variation. However, differently from Wolk et al. (to appear), Szmrecsanyi (2005) provides three different logistic regression models to fit the different datasets corresponding to the three alternations. In his other studies that are more general in scope, Szmrecsanyi has employed quantitative measures to show that the period between 1100 and 1900 does not exhibit a steady drift from synthetic to analytic in English (Szmrecsanyi in press a), and that English is anything but "monolithically analytic" with observable levels of analyticity and syntheticity varying along geographic, short-term diachronic and text type dimensions instead (Szmrecsanyi 2009). Such general takes on the level of analyticity vs. syntheticity of a language are highly desirable and future

research would benefit immensely if the overall level of analyticity vs. syntheticity of an inflectional language like Estonian is studied in detail. The scope of the present thesis does unfortunately not allow such a bird's eye view, but it is nevertheless hoped that the present set of studies makes one of the first contributions.

Following is a short overview of some of the most frequently discussed grammatical alternation phenomena in the literature: the English genitive and dative alternation, particle placement and comparative construction. Focus will be on the type of methodology used (corpus studies vs. experiments), variables included and the results obtained. The reason why I feel it necessary to provide a relatively detailed treatment of these aspects is that the present study has drawn considerable inspiration from these studies – little work has been done on the alternation between synthetic and analytic locative constructions in Estonian and other languages. It will be interesting to see whether an alternation phenomenon that is not an English word order alternation exhibits the effects of similar variables.

The English genitive alternation. From among the long list of previous studies on various alternation phenomena, the English genitive alternation is maybe the one most directly related to the alternation between synthetic and analytic locative constructions in Estonian. Both the English and Estonian alternations involve the parallel use of a synthetic construction (the *s*-genitive in English and the adessive case in Estonian) and an analytic construction (the *of*-genitive and the adposition *peal* 'on'). However, one of the crucial differences between these two alternation phenomena is that while the English genitive alternation is a word order alternation (i.e. in the *s*-genitive construction the possessor comes before the possesum as in *the university's budget*, but in the *of*-genitive construction the possessor follows the possessum as in *the budget of the university*), the Estonian locative alternation is not (i.e. both the adessive case marker *–l* and the adposition *peal* follow the head noun as in *laual* vs. *laua peal* 'on the table'). This characteristic of the alternation between the locative case construction and adpositional construction renders impossible the comparison of some of the major findings about the English genitive alternation. For example, Gries (2002: 22), Hinrichs and Szmrecsanyi (2007: 460), Szmrecsanyi and Hinrichs (2008: 299), Szmrecsanyi (2010: 149, in press b), Shih et al. (to appear) have all found that with long possessors the analytic *of*-genitive is preferred confirming thus the principle of end-weight (Hawkins 1994, 2004; Wasow 1997, 2002). The corpus results presented in Chapter 4 show, however, that for the Estonian locative alternation it is the synthetic adessive case that is preferred with long Landmarks and not the analytic *peal*-construction. Besides length, other variables that have been shown to play a role in the English genitive alternation include the following:

- animacy (Dąbrowska 1998, Gries 2002, Rosenbach 2003, Hinrichs and Szmrecsanyi 2007, Szmrecsanyi and Hinrichs 2008, Szmrecsanyi 2010, in press b, Shih et al. to appear);
- complexity and length (Rosenbach 2005, 2008; Hinrichs and Szmrecsanyi 2007; Szmrecsanyi and Hinrichs 2008; Szmrecsanyi 2010, in press b)
- lexical density (Hinrichs and Szmrecsanyi 2007, Szmrecsanyi and Hinrichs 2008, Szmrecsanyi 2010, Szmrecsanyi in press b);
- thematicity or topicality (Gries 2002, Rosenbach 2003, Hinrichs and Szmrecsanyi 2007, Szmrecsanyi and Hinrichs 2008, Szmrecsanyi 2010, Szmrecsanyi in press, Shih et al. to appear);
- persistence (Hinrichs and Szmrecsanyi 2007, Szmrecsanyi and Hinrichs 2008, Szmrecsanyi 2010, Shih et al. to appear);
- structural parallelism (Hinrichs and Szmrecsanyi 2007);
- the type of possessive relation (Dąbrowska 1998, Rosenbach 2003, Szmrecsanyi in press b);
- phonology (Hinrichs and Szmrecsanyi 2007, Szmrecsanyi and Hinrichs 2008, Szmrecsanyi 2010, Szmrecsanyi in press b, Shih et al. to appear);
- speaker age and gender (Shih et al. to appear).

As to the more specific results pertaining to the English genitive alternation, the synthetic *s*-genitive construction has been found to be preferred with:

- animate possessors (Dąbrowska 1998: 122, Gries 2002: 22, Rosenbach 2003: 398, Rosenbach 2005: 630, Hinrichs and Szmrecsanyi 2007: 464, Szmrecsanyi and Hinrichs 2008: 302, Szmrecsanyi 2010: 152, Szmrecsanyi in press b: 21);
- with lexically dense and more thematic or topical possessors (Rosenbach 2003: 398, Hinrichs and Szmrecsanyi 2007: 460, Szmrecsanyi and Hinrichs 2008: 304, Szmrecsanyi in press b: 21);
- with ownership relations (Dąbrowska 1998: 123, Rosenbach 2003: 398, Szmrecsanyi in press b: 21);
- when the *s*-genitive is preceded by another *s*-genitive (Hinrichs and Szmrecsanyi 2007: 460, Szmrecsanyi and Hinrichs 2008: 304, Szmrecsanyi 2010: 154);
- when the passage where the *s*-genitive occurs is "nounier" than the *of*-genitive passage (Hinrichs and Szmrecsanyi 2007: 460);
- younger subjects tend to use the *s*-genitive form more than older subjects (Shih et al. to appear: 14).

The analytic *of*-genitive, on the other hand, has been found to be preferred with:

- longer possessors (Rosenbach 2005: 630, Hinrichs and Szmrecsanyi 2007: 464, Szmrecsanyi and Hinrichs 2008: 304),

- and with final sibilants in the possessor (Szmrecsanyi and Hinrichs 2008: 303, Szmrecsanyi 2010: 153, Szmrecsanyi in press b: 21, Hinrichs and Shih et al. to appear: 20).

As can be seen from the short overview of the main results found in the literature on the English genitive alternation, there seems to be converging evidence. However, the same studies also provide some interesting conflicting results as to the effect of such variables as givenness or thematicity of the possessor and persistence. Although Rosenbach (2003: 398) and Szmrecsanyi (in press b: 21) report that the *s*-genitive is preferred with more thematic or topical possessors, Gries (2002: 24) and Shih et al. (to appear: 9) do not find a significant effect for thematicity nor givenness in predicting genitive construction choice. Shih et al. (to appear: 10) report that persistence is an insignificant factor in predicting genitive choice, although previous research has found this variable to be significant (Hinrichs and Szmrecsanyi 2007: 460, Szmrecsanyi and Hinrichs 2008: 304, Szmrecsanyi 2010: 154). Furthermore, Szmrecsanyi and Hinrichs (2008: 303) and Szmrecsanyi (2010: 153) report conflicting results as pertains to different registers: thematicity of the possessor is a significant variable in the written data (increased thematicity makes the *s*-genitive more likely), but not in the spoken data. One of the possible explanations for these conflicting results may be the specific method employed in the majority of the papers. Many of these papers use logistic regression to produce the models (and the significant variables) reported to fit the data, where the significance of any one particular variable depends on whatever other variables are included in these models. Given that different researcher have started off with a different set of explanatory variables, at least some amount of divergence in the results is to be only expected.

Other diverging evidence between different registers comes from Gries (2002). He reports (Gries 2002: 23) that while for the written dataset the interaction between the type of genitive construction and length is in the unexpected direction, for the spoken data it is as expected – the possessors are longer than the possessums with the *of*-genitive construction. Gries (2002: 23) considers this result as an important lesson to be learnt for corpus-based analyses of syntactic phenomena – it is crucial to account for the medium or register of the dataset since it is here that processing restrictions like length exert the most influence. Although the author of the present thesis agrees whole-heartedly with this perspective, the corpus studies reported here focus only on written data due to reasons discussed elsewhere in the dissertation (see section 3.1).

However interesting such diverging evidence might be from the perspective of theoretical implications, one has to bear in mind that the differences in the results may very well be due to the type of methodology and principles of operationalization employed by specific researchers or the type of data studied. Gries (2002) and Rosenbach (2003), for example, report the findings from both corpus and experimental studies, while Szmrecsanyi and colleagues restrict themselves to only corpus data. The results also depend on what other variables

have been simultaneously studied and which analytical tools employed. Hinrichs and Szmrecsanyi (2007: 460), for instance, report that the predictor 'givenness of the possessor head' is only selected as significant when length and thematicity of the possessor head are removed from the model (cf. the comment above about regression modelling). The reason why Hinrichs and Szmrecsanyi (2007: 460), Szmrecsanyi and Hinrichs (2008: 304), Szmrecsanyi (2010: 154), and Shih et al. (to appear: 10) report conflicting findings about the significance of persistence in the English genitive alternation may very well be that Shih et al.'s (to appear) measure of persistence is different from that of the other studies.

The English dative alternation. Although Wolk et al. (to appear: 3) claim that "the dative alternation is one of the most extensively studied alternations in the grammar of English", the jury is still out on whether it is the dative or the genitive alternation that has found more coverage in the variationist literature. From the perspective of the present thesis, the genitive alternation has an upper hand since it is concerned with a similar instance of an alternation between synthetic and analytic constructions as the Estonian locative alternation. It might be, however, useful to take a quick look also at some of the more important variables found to influence the choice between the two dative constructions in English. Similarly to the English genitive alternation, the English dative alternation is considered to be a word order alternation (Gries 2003a: 155), as exemplified in 12 and 13 below.

(12) prepositional (*to*-)dative (Bresnan and Ford 2010: 169):
Who gave that wonderful watch to you?

(13) double object construction (Bresnan and Ford 2010: 169):
Who gave you that wonderful watch?

Wolk et al. (to appear), Bresnan and Ford (2010), and Bresnan et al. (2007) report the results of a logistic regression analysis of corpus data. All three studies report the significant effect of the following variables: length, animacy, definiteness, (pro)nominality, and variety. In addition, Wolk et al. (to appear) report the significant effect of register and time; Bresnan et al. (2007) and Bresnan and Ford (2010) of number, verb lemma and sense, and structural parallelism. By employing a mixed-effects logistic regression analysis, both Bresnan and Ford (2010) and Wolk et al. (to appear) convincingly demonstrated the usefulness of this method. Wolk et al. (to appear: 20) report that the random variable of 'verb lemma', i.e. the specific verb used in the two dative constructions, accounts for a large amount of variation. In addition to corpus studies, Bresnan and Ford (2010) also report converging results from a series of psycholinguistic experiments: rating the naturalness of alternative dative constructions, continuous lexical decision task, and sentence completion.

Particle placement in English. Together with the English genitive and dative alternation, particle placement in English is another alternation phenomenon that has found extensive coverage in the literature. The difference between the two constructions lies in where the particle is placed – it either precedes the direct object as in *John picked up the book* or follows it as in *John picked the book up*. Most relevant to the study on the Estonian locative alternation are the studies carried out by Gries (1999, 2001, 2002, 2003a, 2003b) since he is known to be an avid proponent of the multivariate approach taken also in the present thesis. Gries (2003a: 165) reports that the following variables discriminate between the two constructions: length, pronominality, complexity, definiteness, concreteness of the direct object, idiomaticity of the verb, structural parallelism (times and distance of last mention). Animacy which is given as one of the main variables in the English genitive and dative alternation does not seem to play a role in Gries's (2003a) dataset. Although Gries (2003a: 161) lists the phonological variable of stress pattern as one of the variables argued to contribute to particle placement, he does not test the relevance of this variable in this particular dataset. He has, however, touched upon this issue in his other work (e.g. Gries 1999), where it is shown that when the direct object is stressed, it tends to follow the particle.

Other relevant alternation phenomena. The other relevant alternation phenomena include interrogative clause linkers in English (Rohdenburg 2003), and the English comparative construction (Mondorf 2003). Rohdenburg (2003) presents the results of various case studies concerning the linkers in interrogative clauses in English and discusses the implications of the complexity principle and the *horror aequi* principle. Although the corpus methodology used in Rohdenburg's (2003) analysis is not as sophisticated as the logistic regression models reported in, for example, Hinrichs and Szmrecsanyi (2007), Szmrecsanyi and Hinrichs (2008), Bresnan and Ford (2010), Szmrecsanyi (2010), Szmrecsanyi (in press b), and Wolk et al. (to appear), it does provide valuable theoretical discussion on the issues of complexity. Nevertheless, Rohdenburg's (2003) study is an instance of when the potentially significant qualitative results fail to be supported by strong and convincing quantitative results. Mondorf's (2003) study of the English comparative construction is relevant to the present study in that it studies an alternation between a synthetic (*prouder*) and analytic construction (*more proud*). However, it should be once again noted that similarly to all the previous alternation phenomena described above – the English genitive and dative alternation and the particle placement – the English comparative constructions also exhibit word order variation. In the synthetic variant, the comparative marker follows the head, but in the analytic variant it precedes the head (cf. the adessive case marker and the adposition *peal* which both follow the head). The main argument of Mondorf's (2003: 252) study is that the analytic variant is resorted to "whenever a structure requires more processing capacity, be it for matters of phonology, morphology, lexicon, syntax, semantics or pragmatics". However, the data presented comes only from

a corpus and suffers similar drawbacks as Rohdenburg's (2003) study as pertains to the balance between qualitative and quantitative results with the former presenting a strong argument, but not sufficiently supported by the latter.

2.2.2 Summary of previous alternation studies

Table 6 provides in a summary fashion an overview of the various variables (given in alphabetical order) found to play a role in the various alternation phenomena in English described in the previous section. Although one should be careful in drawing any far reaching conclusion from this simplified table based on a limited number of alternation studies, it can be seen that there are a number of variables that have been found to play a role in the majority of the alternations discussed: animacy, length, and structural parallelism. There seems to be evidence and justification for including similar variables in the analysis of the alternation between Estonian locative cases and adpositions. Yet, as mentioned above, the alternation phenomena covered do not elicit direct comparisons with the Estonian locative alternation because unlike the English alternations, the Estonian alternation is not a word order alternation. In word order variations, the variables animacy and length or weight are reported to constitute some of the major constraints on the choice of construction; their effects seem to derive from cognitive or processing constraints (e.g. Bock et al. 1992, McDonald et al. 1993, Hawkins 1994, Wasow 2002 inter alia).

Table 6. Overview of significant variables in English alternation phenomena

Variable name	Genitive altern.	Dative altern.	Particle placement
Animacy	✓	✓	✓
Complexity			✓
Concreteness			✓
Definiteness		✓	✓
Idiomaticity			✓
Language variety		✓	
Length	✓	✓	✓
Lexical density	✓		
Number		✓	
Persistence	✓		
Phonology	✓		
Pronominality		✓	✓
Register	✓	✓	
Speaker age	✓		
Speaker gender	✓		
Structural parallelism	✓	✓	✓
Thematicity or topicality	✓		
Time	✓	✓	
Type of relation	✓		
Verb lemma		✓	

An important aspect that renders the direct comparison of the results found in the various alternation studies difficult is the fact that all of the studies reviewed pertain only to the English language. It is not the aim of the this overview to leave the reader with the impression that alternations do not exist in other languages or that researchers do not study them, but unfortunately this simply reflects the status-quo of today's linguistic research where the vast majority of the publications in first-class international journals predominantly deal with the English language. This bias towards the English language is especially prominent in the research domain of alternations. One of the reasons why specifically these alternations in the English language have been studied in detail is that "pairs of semantically more-or-less equivalent expressions [...] have captured the attention and imagination of researchers working in many different theoretical paradigms" (Gries and Stefanowitsch 2004b: 97). That is to say these specific alternations provide excellent food for thought and argumentation for linguists concerned with proving the validity of their preferred theory.

The type of analysis employed in many of the studies reviewed in the previous sections, i.e. the logistic regression analysis, allows also a ranking of the significant variables. Hinrichs and Szmrecsanyi (2007: 464) for example report that the most important variable in their study on the English genitive alternation is animacy of the possessor, followed by length of the possessor. On the whole, they propose the following hierarchy of relevance:

> semantics/pragmatics ~ processing/parsing > phonology > economy (Hinrichs and Szmrecsanyi 2007: 464)

Rosenbach, on the other hand, arrives at the following ranking of variables in her study of the English genitive alternation:

> animacy > topicality > possessive relation (Rosenbach 2003: 398)

The two hierarchies do not allow, however, direct comparison since Rosenbach's (2003) study only included these three variables, while Hinrichs and Szmrecsanyi (2007) considered a variety of different variables. Nevertheless, there seems to be no doubt that animacy is the most important variable in the choice between the two genitive constructions in English as both Gries (2002: 20) and Shih et al. (to appear: 13) also provide converging evidence. As to the dative alternation, Bresnan and Ford (2010: 182) report that the most important variable is the pronominality of the recipient and theme, followed by length and definiteness. For the particle placement, Gries (2003a: 166) concludes that the variable group that is the most decisive for the choice of construction is "without exception" morphosyntactic and there seems to be no question about the superiority of morphosyntax over discourse-pragmatics.

Hinrichs and Szmrecsanyi (2007: 464) conclude for their study on the English genitive that animacy and end-weight as pillars of their logistic regres-

sion model provide empirical support to Rosenbach's claim that animacy effect "cannot be reduced to an effect of weight (and vice versa)" (2005: 638). Rosenbach (2005, 2008) discusses whether animacy effects are simply an artefact of other variables, such as, for example, syntactic weight, or whether animacy is an independent variable in grammatical variation. She concludes that animacy is not an epiphenomenon of other factors in grammatical variation, despite its statistical correlation with variables such as topicality and weight. She makes reference to the following studies as converging evidence: Rosenbach (2002, 2005) and Bresnan et al. (2007). According to Rosenbach's (2005: 621) experimental results animacy and weight are independent factors and neither can be reduced to the other. Rosenbach (2005) reports the findings of a forced-choice task where the variables animacy and weight were controlled for separately. Complexity and length is another pair of variables that correlate considerably with each other – longer constituents tend to be more complex and vice versa (Rosenbach 2005: 617). Wasow and Arnold (2003: 120–128) provide a nice overview of the controversy regarding whether length and complexity are distinct variables or not.

The vast majority of the studies on the various alternation phenomena reviewed for the present thesis employ corpus methodology. The proportion of experimental work reported in these studies is relatively small. Gries (1999, 2002), Rosenbach (2003, 2005), Bresnan and Ford (2010) provide alternatives by employing an acceptability task, a forced choice task and a combination of the two respectively. Although Gries (2002: 26) reports conflicting findings as to the results found in previous studies on the English genitive alternation, he does provide converging evidence between the different methodologies he employed – the corpus data correspond to the experimental acceptability judgement data. Rosenbach (2005) provides data from a corpus study as well as the results of an experimental study (a forced choice task), showing that animacy and weight are independent factors and neither can be reduced to the other. She (Rosenbach 2005) furthermore argues that animacy is a processing factor influencing grammatical variation, just as weight is.

Bresnan and Ford (2010) report the findings of three experiments on the dative constructions in American and Australian varieties of English; these experimental results are compared to the probabilistic models of corpus data. The experiments Bresnan and Ford (2010) employ are the following: a sentence rating task, continuous lexical decision, and sentence completion. Relevant from the perspective of the dissertation are the first and the last experiment. In the sentence rating task (Bresnan and Ford 2010: 184–191), the participants were asked to rate the relative naturalness of both of the given dative alternatives with 100 points to express their rating so that the ratings for any pair of alternatives added up to 100. Arppe and Järvikiv (2007: 150–151) have criticised this kind of technique of putting together in essence two different types of experiments (e.g. a forced choice task and a judgement task) as problematic. Participants are likely to perform two different activities for these two tasks and when interpreting the results, it is difficult to tease apart

production from comprehension. The second relevant experiment employed by Bresnan and Ford (2010: 199–201) is a sentence completion task where participants were asked to complete an unfinished sentence following a text passage whichever way felt most natural to them. This kind of production task is a valuable method for studying grammatical synonymy and one possible line of future research on the alternation between Estonian locative cases and the corresponding adpositions is to employ a similar experimental design. The results of such an experiment could then be compared to the actual corpus results to see how much the results converge or diverge.

2.3 Variable selection

This section introduces the semantic and morphosyntactic variables included in the studies along with specific examples and a detailed discussion of how the variables were operationalized. The variables can be broadly divided into two groups: semantic variables (section 2.3.1) and morphosyntactic variables (section 2.3.2). Within the group of semantic variables, the type of relation between the Landmark and Trajector is presented first, followed by the different properties of both Landmark and Trajector (type, animacy, number, and mobility) and the relative size of the Trajector in relation to the Landmark. The last semantic variable described is the verb lemma. This ordering of variables proceeds from the logic of first identifying the type of relation between the two entities, then moving on to characterising the two entities separately, and concluding with a different category – that of the verbal predicate. Within the morphosyntactic variables, the different formal properties of the Landmark phrase are presented first – length, morphological complexity, and syntactic function. Length and complexity are interrelated and the discussion starts with these variables, because previous studies have shown that this set of variables plays a prominent role in alternation phenomena. After the morphosyntactic properties of the Landmark phrase, the variables word class of both Landmark and Trajector and the case form of Trajector are presented. The final three morphosyntactic variables look at a larger context than the level of the Landmark and Trajector phrase – clause type and word order.

The majority of the variables were selected using the findings from studies on similar phenomena in Finno-Ugric languages, studies on different alternation phenomena in other languages, other relevant literature described in detail in the previous section, and my own native-speaker intuition. I also discuss in section 2.3.3 the variables which were not considered in the present thesis, but which were, however, shown to play a role in other alternation studies, and the reasons for excluding them for the time being.

The examples used in illustrating the different semantic and morphosyntactic variables are taken from two sources: the fiction and newspaper subcorpora of the Morphologically Disambiguated Corpus of Estonian (MDCE) and the Balanced Corpus of Estonian (BCE); see section 3.1 for details. The abbrevia-

tions used at the end of each example sentence, MDCE and BCE respectively, refer to the source of the example. The exact spelling and form of the example sentences as found in the corpus is retained, with the following exceptions: the spaces before the punctuation marks have been removed and some very long examples sentences have been abbreviated when possible with the symbol [...] specifying the excluded part.

2.3.1 Semantic variables

Numerous cognitive-functional studies on spatial language expressions have shown that various properties of Trajectors and Landmarks participating in the locative constructions influence the use of spatial expressions (e.g. Talmy 1983, Herskovits 1986, Vandeloise 1991, Feist and Gentner 2003, Coventry and Garrod 2004, Carlson and Van der Zee 2005). In line with this research tradition, the following 11 semantic predictors were included in the present study: type of relation between Trajector and Landmark; type, animacy, number, and mobility of Landmark and Trajector; relative sizes of Trajector and Landmark; and type of verb. It should be noted that coding the datasets for semantic variables involves considerably more subjective decisions than the coding for formal variables. By taking previous studies as examples and adopting a rigorous coding schema it is hoped that at least some of this subjectivity has been eliminated.

Type of relation between Landmark and Trajector. Both the Estonian adessive and the adposition *peal* 'on' can express spatial and abstract relations between a Trajector and a Landmark. It has been suggested in previous work on cases and adpositional constructions that cases are semantically more abstract than adpositions (Bartens 1978, Comrie 1986, Erelt et al. 2007: 191, Luraghi 1991: 66-67, Ojutkangas 2008, Hagège 2010: 37-38, Lestrade 2010). Hence, it is feasible to posit that the case construction is the preferred construction when the relation between a Landmark and a Trajector is abstract. It is undoubtedly true that the adessive case fulfils many grammatical or abstract functions (e.g. expressing the agent or the possessor), as was seen above in section 2.1.1; in these functions, the adpositional construction is not a possible alternative and all of these instances were excluded, as described in detail in section 3.1. However, there are also instances of abstract relations between a Landmark and a Trajector, where both the synthetic and analytic constructions can be used. Such abstract relations mostly involve meaning transfer, as *hingel* and *hinge peal* 'on soul' in examples 14 and 15 below:

(14) [...] *mida* *juba* *aastaid* *nagu* *soomussärki* ***oma***
what.SG.PRT already year.PL.PRT like bullet-proof shirt.SG.PRT my
tegelikult ***nii*** ***kergesti*** ***haavatavал*** ***hingel***
actually so easily wounded.SG.ADE soul.SG.ADE
kannan? <MDCE: fiction>
wear-PRS.1SG
'[...] that I have been wearing **on my so easily wounded soul** for years like a bullet-proof shirt?'

(15) *Nagu* *inimene,* *kellel* *on* *midagi*
like person.SG.NOM who.SG.ADE be-PRS.3SG something.SG.NOM
hinge ***peal***. <BCE: fiction>
soul.SG.GEN on
'Like a person who has something **on his soul**.'

Furthermore, Bartens (1978) demonstrates that in the Saami languages analytic constructions are used when the spatial relation between Trajector and Landmark is noncanonical, i.e. it differs to some extent from the everyday situation. In addition, Luraghi (1991: 60) has pointed out that noncanonical spatial relations are not "normal" and they need extra-marking, which in Estonian can be done by using the adposition *peal* 'on'. Klavan et al. (2011) report the results of a forced choice task and a production task which, among other things, tested whether the type of relation between Landmark and Trajector plays a role in the use of the Estonian adessive and adposition *peal* 'on'. The results indicate that when there is an unconventional spatial relation between Trajector and Landmark (e.g. a book placed on top of an alarm-clock), the adpositional construction tends to be used and when the relation is abstract, the case construction is preferred. However, since the experimental designs described in Klavan et al. (2011) were not balanced and the analysis of the results does not generalise to other stimuli and subjects, it was decided to study the issue of abstract vs. spatial relations further in the present thesis in the multivariate corpus study. This variable has two levels: 'abstract' and 'spatial'.

Type of Landmark. It can also be predicted that there is a general difference between what types of Landmarks are used together with either the locative cases or adpositions. A general distinction is made between small easily manipulable objects or 'things' (e.g. *kelk* 'sleigh' in example 16) and large static objects or 'places' (e.g. *kallas* 'shore' in example 17):

(16) *Kotipundar* ***kelgu*** ***peal*** *vabises.* <BCE: fiction>
bundle of bags.SG.NOM sleigh.SG.GEN on quiver-PST.3SG
'The bundle of bags **on the sleigh** quivered.'

(17) Ma seisin õnnelikult pääsenuna **kaldal** [...] <MDCE: fiction>
I.NOM stand-PST.1SG happily survivor.SG.ESS shore.SG.ADE
'I was standing as a happy survivor **on the shore** [...]'

Bartens (1978) demonstrated that in the Saami languages the synthetic constructions are used when the Landmark is a place and analytic constructions are more frequent with things as Landmarks. The same has been demonstrated by Ojutkangas (2008), who studied the use of the interior locative cases and the corresponding adpositions in Finnish. Although Ojutkangas (2008) does not claim so, it can be concluded, based on the data she presents, that the case construction is more frequent with body parts as Landmarks. Thus there seems to be ample reason to suspect that the type of Landmark plays a role in the alternation between locative cases and adpositions. More specifically – larger locations such as places should lend themselves more easily for abstraction and hence are more likely to be used with the adessive (cf. e.g. Bartens 1978), while as small manipulable objects or things prefer the adposition *peal* 'on' since adpositions are more concrete and specific than cases and they convey the meaning of spatial location of an object more clearly (Bartens 1978, Palmeos 1985: 18, Comrie 1986, Luraghi 1991: 66-67, Ojutkangas 2008, Hagège 2010: 37-38, Lestrade 2010).

Type of Trajector. Although the focus in the present study is on Landmarks – this is where the linguistic phenomenon under study exhibits itself (i.e. the marking of the Landmark either with the locative case or the adposition) – Trajectors are also important components of spatial expressions and should not be discarded. In fact, as Langacker (2008: 70) points out, Trajector is the most prominent participant when a relationship is profiled. Naturally, human beings are interested in the entity that is located and it is plausible that certain types of Trajectors predict the use of the synthetic construction and other types of Trajectors the analytic construction. This variable has two levels: 'abstract' and 'object'. When the Trajector was not overtly expressed in the sentence, a larger context was used to determine the semantic type of Trajectors. Abstract Trajectors frequently involve abstract concepts like emotions or instances as specified in example 18 – *see kevadine vahejuhtum Tartus* 'this spring incident in Tartu'.

(18)	[...] *tal*	*kripeldas*	***see***	***kevadine***	***vahejuhtum***
	he.ADE	prickle-PRS.3SG	this.SG.NOM	spring.SG.NOM	incident.SG.NOM
	Tartus	*ikka veel*	*hinge*	*peal.* <BCE: fiction>	
	Tartu.INE	still	soul.SG.GEN	on	

'[...]he still had **this spring incident in Tartu** prickling his soul.'

Trajectors coded as 'objects' may refer to animate objects (i.e. to humans or animals) or to inanimate objects, as *fotod* 'photos' in example 19. As is often the case with Trajectors, they can express activities or whole events (Zlatev 2007: 327), as *ta laulab* 'he is singing' in example 20 below. In such cases, it was decided to look at the whole context of a sentence and to code the type of Trajector according to what was actually located on top of the Landmark. In

example 20, the Trajector is the personal pronoun *ta* 'he' and it was coded as 'animate'.

(19) *Linnagaleriis* *on* *väljas* *kaks* *valget*
town.gallery.SG.INE be-PRS.3PL out two.SG.NOM white.SG.NOM
metallkasti *miniatuursete* *kõlaritega,* *mille*
metal.box.SG.PRT miniature.PL.COM loudspeaker.PL.COM what.SG.GEN
peal ***fotod***. <BCE: newspapers>
on photo.PL.NOM
'There are two white metal boxes with miniature loudspeakers, which have **photos** on top, exhibited in the town gallery.'

(20) ***Ta*** ***laulab*** *voodi* *peal* . <BCE: newspapers>
he.NOM sing-PRS.3SG bed.SG.GEN on
'**He is singing** on the bed.'

Animacy of Landmark. Since animacy is considered a very important cognitive category and is discussed in numerous linguistic and psycholinguistic studies (for overviews, see for example, de Vega et al. 2002: 121–122, Feist and Gentner 2003: 2, Rosenbach 2005, 2008, Bresnan and Ford 2010: 10), it was decided to code the Estonian adessive and adposition *peal* 'on' dataset for this category as well. Animacy is considered a very influential variable, if not the most important variable, in determining the choice between the *s*-genitives and *of*-genitives in English (e.g. Rosenbach 2003, 2005, 2008; Hinrichs and Szmrecsanyi 2007; Wolk et al. to appear). Rosenbach (2008: 151) claims that if there are two ways of expressing the same thing, it is animacy that in many cases determines the choice between alternative constructions. Since the English genitive alternation is a word order alternation, it makes sense to tie this alternation to the general finding in the literature about animacy and linear order – there is psycholinguistic evidence (Bock et al. 1992, McDonald et al. 1993) that shows the preference for animate referents to be placed first in linear order in English. It is therefore claimed that the *s*-genitive prefers animate possessors and there exists a large body of converging evidence (e.g. Rosenbach 2003, 2005, 2008; Hinrichs and Szmrecsanyi 2007; Wolk et al. to appear,). It is also continuously pointed out in the literature on the English genitive alternation that animacy is a variable that is highly correlated with other variables such as topicality, definiteness and syntactic weight or length. According to Rosenbach (2008: 156) animates, especially humans, are more likely to be discourse topics, definite, and shorter than inanimates. However, she (Rosenbach 2005, 2008) has fervently argued and presented evidence that animacy is a genuine factor in grammatical variation and not just something that is epiphenomenal, i.e. something that just correlates with other more important and real factors like weight (cf. Hawkins 1994, 2004).

Although Rosenbach (2008: 157) claims that the results pertaining to animacy and other variables shown to influence the English genitive variation transfer easily to other languages and to other cases of grammatical variation in

English, reference is only made to studies on Modern Low Saxon (a Germanic language like English) and to the dative alternation in English. Hence, it was decided to see if animacy plays a role in the alternation between Estonian locative constructions as well. In discussing the animacy of the Landmark, we have to come back to the issue that has already been mentioned on several occasions – the fact that the Estonian adessive case fulfils many other functions besides expressing location. For instance, it can express the logical subject in possessive constructions and it is typically humans who possess things. Therefore, in order to avoid ambiguity, it can be predicted that when there is indeed a need to express physical location of a Trajector on top of an animate Landmark, the preferred construction is the adpositional construction.

Following, for example, Bresnan et al. (2007) and Shih et al. (to appear) the operationalization of animacy was based on the simple distinction between 'animate' and 'inanimate' categories. Animate Landmarks comprise humans, higher animals and human-like beings such as gods; *hobune* 'horse' in (21) is an example of an animate Landmark. All other Landmarks were classified as inanimate. I agree with Rosenbach (2005: 623 Ftn. 22) and others who point out that it is often difficult to decide whether collective nouns, e.g. lexemes denoting official bodies and organizations are animate or inanimate. In general, I decided to classify them as animate; they were considered inanimate only if they did not involve agency.

(21) [...] *ega* *tahtnud* *pikal* *teel* ***hobuse*** ***peal***
nor want-PST.3SG long.SG.ADE road.SG.ADE horse.SG.GEN on
istudes *külmetada* . <BCE: fiction>
sit-PRS.PTCP freeze-INF
'[...] and didn't want to freeze herself sitting **on a horse** on a long road.'

Animacy of Trajector. As with the operationalization of animacy of the Landmark, the animacy of the Trajector was based on the simplified distinction between 'animate' and 'inanimate' categories. The same criteria apply as above. It would be interesting to see if and how the animacy variable correlates with other variables mentioned in the literature, e.g. topicality, weight and linear order. I will come back to this issue when I discuss these other variables. I will only mention that, based on the evidence presented in literature on other variation phenomena, it can be predicted that there is a correlation between animacy and the word order between Trajector and Landmark phrases: when the Trajector is animate and the Landmark inanimate, the preferred word order is to place the Trajector first before the verb and the Landmark. Trajectors are also considered to be more topical than Landmarks.

Number of Landmark. Another cognitively and typologically important category in grammar is number (Greenberg 1966: 27–39). Greenberg (1996) discusses the concept of marked and unmarked categories and how this relates to phonological, grammatical, and semantic aspects of language. Rohdenburg

(2003: 223) has pointed out that "[p]roponents of markedness theory generally assume that the morphologically marked plural noun represents a cognitively more complex category" and that "[i]t may be hypothesised, therefore, that constructions associated with plural nouns tend to exhibit a greater degree of grammatical explicitness than their singular counterparts". Number has been shown to play a role in the dative alternation, where plural themes prefer the prepositional dative (Bresnan and Ford 2010: 176–179). Based on these claims it can be predicted in relation to Estonian locative cases and adpositions, that since plural nouns show a greater degree of explicitness, they do not require an adposition like *peal* to mark extra-explicitness; instead, they make do with the more concise adessive case. Landmarks were coded either 'plural' or 'singular' in both corpus studies. When context and formal plural marking were in conflict or when there was ambiguity, the analysis proceeded from the context and relied on the semantic coding. This is the reason why this variable is considered under the category of semantic variables and not morphosyntactic variables.

Number of Trajector. As with the operationalization of number of the Landmark, the number of the Trajector was based on the distinction between 'plural' and 'singular' categories. The same criteria and theoretical reasoning apply as above.

Mobility of Landmark. Following de Vega et al. (2002), mobility was another property of Landmarks and Trajectors that was studied in the datasets of the Estonian adessive and adposition *peal* 'on'. Landmarks were coded as either 'mobile' or 'static'. Mobile Landmarks are those that do not have a fixed position in the environment, either because they move by themselves (e.g. humans, animals) or can be moved by an external agent (e.g. a table). Static Landmarks (the majority of which in the dataset are also places, but not all) have a fixed position in the environment (e.g. street, market). Example 22 illustrates an instance where the Landmark, *kušeti äär* 'the edge of the couch', is coded as 'static':

(22)	*Ruudi*	*istub*	***kušeti***	***ääre***	*peal* [...]
	Ruudi.SG.NOM	sit-PRS.3SG	couch.SG.GEN	edge.SG.GEN	on
	<BCE: fiction>				
	'Ruudi is sitting on **the edge of the couch** [...]'				

Mobility of Trajector. Similarly to mobile Landmarks, mobile Trajectors are those that do not have a fixed position in the environment, either because they move by themselves or can be moved by an external agent, as *koer* 'dog' in example 23. Static Trajectors have a fixed position in the environment, as *jäljed* 'tracks' in example 24.

(23) [...] ***koer*** *ei* *tohi* *magada* *kajutis* *voodis,*
dog.SG.NOM no can-PRS.3SG sleep.INF cabin.SG.INE bed.SG.INE
diivanil *ega* *tooli* *peal.* <BCE: newspapers>
couch.SG.ADE nor chair.SG.GEN on
'[...] **a dog** must not sleep in bed, on the couch or on the chair in the cabin.'

(24) ***Jäljed*** *olid* *lume* *peal* *hästi* *näha.*
track.PL.NOM be-PST.3PL snow.SG.GEN on well visible
<BCE: newspapers>
'**The tracks** were clearly visible on the snow.'

Relative size of Trajector in relation to Landmark. In cognitive linguistic analyses of spatial expressions, it has been claimed that Landmarks tend to be larger than Trajectors (e.g. Talmy 1983: 230–231, 2000: 315–316). In order to validate this claim and to see whether this variable influences the use of the Estonian adessive case and the postposition *peal* 'on', the relative size of the Trajector in relation to the Landmark is judged either as 'conventional', 'same' or 'unconventional'. The relative size of the Trajector in relation to the Landmark is conventional when the Landmark is bigger than the Trajector, as in example 25 where the Landmark *katus* 'roof' is judged to be bigger than the Trajector *Miša* (a proper name referring to a cat); the relative size is considered 'same' when the Trajector and Landmark are more or less of the same size, as *Mihkel* (a proper name referring to a male person) and *sohva* 'couch' in example 26; finally, the relative size of the Trajector in relation to the Landmark is unconventional when the Trajector is bigger than the Landmark, as in example 27 where *mees* 'a man' is judged to be bigger than either *pliit* 'stove' or *taburet* 'tabouret'.

(25) ***Miša*** *on* ***katusel.*** <MDCE: fiction>
Miša.SG.NOM be-PRS.3SG roof.SG.ADE
'**Misha** is **on the roof**.'

(26) ***Mihkel*** *aga* *oigas* *veel* *tükk* *aega* ***sohva*** ***peal,***
Mihkel.SG.NOM but whine-PST.3SG still some time couch.SG.GEN on
enne kui *valu* *pisut* *järele andis* [...] <BCE: fiction>
before pain.SG.NOM a little alleviate-PST.3SG
'**Mihkel** was whining **on the couch** for some time, before the pain alleviated [...]'

(27) ***Pliidi*** ***peal*** ***tabouretil*** [...] *istub* *paksu*
stove.SG.GEN on tabouret.SG.ADE sit-PRS.3SG thick.SG.ILL
kasukasse *mässitud* ***mees*** . <BCE: fiction>
fur-coat.SG.ILL wrapped.SG.NOM man.SG.NOM
'**A man** wrapped in a thick fur-coat is sitting **on the stove on a tabouret** [...]'

Since the dataset of the corpus study is comprised of written texts only, without visual representation of the situation described, the evaluation given as to the relative sizes of the Landmark and the Trajector is considerably more subjective than the coding of other variables. However, I have relied on the standard sizes of the various objects denoted by the Trajector and the Landmark. Moreover, this predictor is closely related to the predictors 'type of Landmark' and 'type of Trajector'. For example, with things as Landmarks, Trajectors are conventionally of the same size or sometimes even larger than Landmarks (examples 26 and 27), but with places as Landmark, Trajectors tend to be smaller than Landmarks. Since adpositions are perceived in the literature as expressing more specific locations than cases, it is predicted that when the relative size of the Trajector in relation to the Landmark is unconventional, i.e. the Trajector is bigger than the Landmark, the adpositional construction is preferred.

Verb lemma. The verbal component of locative expressions has been a relatively neglected topic in linguistics (Ameka and Levinson 2007: 847). However, the volume of collected papers edited by Newman (2002), the special issue of *Linguistics* (2007, vol. 45-5), and studies by researchers like Newman and Rice (2004), Lemmens (2002, 2005, 2006), Lemmens and Perrez (2010) and others show that the locative predicate plays an important role as well. The majority of the listed papers focus on posture verbs, i.e. verbs like *sit, stand,* and *lie*, and emphasise how in some languages the properties of both the Trajector and Landmark or other relevant contextual information require the use of a specific posture verb in locative expressions. Estonian is similar to English in that it has the potential of using different locative verbs, including posture verbs, but it still prefers the use of a simple copula to express location. However, such claims need further corroboration, since the verbal component of locative expressions has not been studied in Estonian linguistics. One of the minor aims of the multivariate corpus study reported in the dissertation is to provide an overview of what type of verb lemmas are used in expressing support-relations. Nonetheless, care should be taken in drawing any far-reaching conclusions about the verbal component in Estonian locative expressions in general, since I am looking at two very specific constructions – the adessive and the adposition *peal* 'on'.

As pointed out, the predominant locative construction in Estonian tends to use a simple copula in expressing location (e.g. *on* 'are' in example 28), but there are also other verbs that can be used together with either the adessive or the adposition *peal* 'on' (e.g. *istus* 'was sitting' in example 29).

(28) *Selle pildiga, kus munad **on** voodi **peal**,*
this picture.SG.COM where egg.PL.NOM be-PRS.3PL bed.SG.GEN on
oli selline lugu. <BCE: newspapers>
be-PST.3SG this story.SG.NOM
'With this picture, where **the eggs are** on the bed, was the following story.'

(29) [...] *kui* *ma* *Miia* *juurde* *läksin,* ***istus***
when I.SG.NOM Miia.SG.GEN to go-PST.1SG sit-PST.3SG
ta ***diivanil*** […] <MDCE: fiction>
she.NOM couch.SG.ADE
'[...] when I went to Miia, she **was sitting on the couch** [...]'

In the initial coding stage, each sentence in the two datasets was coded for the verb lemma used with either the adessive or the *peal*-construction. In the next stage, the verbs were subcategorised into different groups based largely on Levin (1993) and include the following groups: 'action verbs' (e.g. *avama* 'open'), 'existence verbs' (e.g. *olema* 'be'), 'motion verbs' (e.g. *jooksma* 'run'), and 'posture verbs' (e.g. *istuma* 'sit'). In addition, this variable also had the level of 'no verb' – this is used for elliptical sentences where no overt verb lemma was expressed, as in example 30.

(30) *Toit* ***laual*** – *nõud* *küürida* [...] <MDCE: fiction>
food.SG.NOM table.SG.ADE dishes.NOM scrub-INF
'Food **on the table** – the dishes to scrub [...]'

The simple copula *olema* 'be' is semantically empty, i.e. it does not say anything about the properties of the Trajectors or Landmarks nor does it specify the exact type of location (e.g. sitting vs. lying vs. standing on something), and the same verb is also used in the possessive construction with the adessive case. It is therefore predicted that with the simple copula *olema* 'be', the more specific adpositional construction is preferred, while other types of verbs prefer the more abstract and less specific case construction. This line of reasoning is based on the assumption that language speakers wish to avoid the potential ambiguity between the locative and possessive functions of the adessive case and prefer to use the adpositional construction with the verb *olema* 'to be'. In addition, it can be argued that because the adessive case fulfils many other functions besides expressing location, its meaning is more bleached vis-à-vis its locative meaning than that of the adposition *peal* 'on'. Consequently, speakers may very well compensate for the "emptiness" or the low degree of semantic specificity of the locative construction (i.e. the adessive case construction) by making the verb in the spatial expression carry more meaning.

2.3.2 Morphological variables

This section describes the morphosyntactic variables considered in the corpus study. Morphosyntactic variables have to do with formal aspects of the Trajector and Landmark phrases and various properties of the clause in which the two constructions occur. This group includes the following nine variables: length of the Landmark phrase, morphological complexity of Landmark, syntactic function of the Landmark phrase, word class of Landmark and Trajector, case form of Trajector, clause type, position of the Landmark phrase

in the clause, and the relative position of the Trajector and Landmark. In the variationist literature, the morphosyntactic variable of length or complexity (also referred to as weight) has been found to play a prominent role (Mondorf 2003; Rohdenburg 2003; Rosenbach 2003, 2005; Bresnan et al. 2007; Hinrichs and Szmrecsanyi 2007; Szmrecsanyi and Hinrichs 2008; Szmrecsanyi 2010, Szmrecsanyi in press b; Shih et al. to appear).

Length of the Landmark phrase. Length is one of the most crucial variables in numerous studies on various syntactic alternation phenomena (e.g. Cooper and Ross 1975, Hawkins 1994, Wasow 1997, Arnold et al. 2000, Wasow 2002, Wasow and Arnold 2003, Hawkins 2004, Bresnan et al. 2007, Anttila et al. 2010, Shih et al. to appear). In many cases length is discussed under such headings as weight or complexity; more specifically reference is made to the "time-honoured principle of 'end-weight'" (Hinrichs and Szmrecsanyi 2007: 453) and "Rohdenburg's complexity principle" (Mondorf 2003: 294).

The **principle of end-weight** expresses the idea that short elements should precede long or 'heavier' ones and that phrases are ordered with increasing weight (Wasow 2002: 3). According to Rosenbach (2005), this principle is closely related to Hawkins's (1994, 2004) parsing theory – the parser wants to process information as efficiently as possible and placing 'heavier' constituents after shorter ones facilitates parsing. As such, the principle of end-weight is closely related to word order and has been found to play a significant role in syntactic word order alternations, the English genitive and dative alternations and the particle placement in English verb-particle constructions being excellent examples. For instance, the *s*-genitive has been shown to be more frequent with short possessors (Rosenbach 2005: 614, Shih et al. to appear), and the Verb-Particle-Object construction is strongly preferred with long direct objects (Gries 1999: 110). However, it should be born in mind that unlike the English genitive and dative alternations as well as the phenomenon of particle placement, the alternation between the synthetic locative cases and analytic adpositional constructions in Estonian is not a word order alternation. Nevertheless, it is still presumed that there is an effect for length, albeit with different consequences. I return to this specific prediction below.

It is not always clear weather the term 'weight' refers to 'length' or to 'complexity' or to both. Rosenbach (2005: 617) reports that Altenberg (1982 cited in Rosenbach 2005) considers weight as length (i.e. number of words), but that others use it to refer to complexity or for length and complexity together (i.e. structural complexity, e.g. Wasow 2002, Wasow and Arnold 2003: 120-128). It is certainly true that complexity and length are correlated and that longer constituents tend to be more complex and vice versa (Rosenbach 2005: 617). However, since it is still a matter of controversy whether they are two distinct factors or not (Rosenbach 2005: 617), the two sides of 'weight' are kept separate in this study – both the length of the Landmark phrase as well as the syntactic complexity of the Landmark phrase are included as two separate variables.

The other principle that is frequently discussed in alternation phenomena and that is also related to the variable 'length' is **Rohdenburg's complexity principle** (Mondorf 2003: 294, Rohdenburg 2003). The principle states that "in the case of more or less explicit constructional alternatives, the more explicit option(s) will tend to be preferred in cognitively more complex environments" (Rohdenburg 2003: 205). In discussing the complexity principle, Rohdenburg (2003: 205) states that he cannot yet support his principle with experiments of his own, but he does make reference to "a vast body of broadly psycholinguistic and typological research" related to concepts like syntactic weight (cf. above), referent accessibility (e.g. Ariel 1990; Gundel et al. 1993 cited in Rohdenburg 2003: 205) and markedness (e.g. Croft 1990; Givón 1991; Wurzel 1998 cited in Rohdenburg 2003: 205). Similarly to Rohdenburg (2003), Mondorf (2003) also takes a general stance on complexity and discusses the complexity principle in her study on English comparative constructions. She (Mondorf 2003) demonstrates how processing capacity, be it due to phonology, morphology, lexicon, syntax, semantics or pragmatics, plays a role in the choice between the synthetic and analytic comparative construction. According to her, "[l]anguage users, when faced with the option between the synthetic and analytic variant, prefer the latter in environments that are for some reason more difficult, more complex, less entrenched, less frequent, less accessible or in any way cognitively more complex." (Mondorf 2003: 252). Mondorf (2003) argues for a presumably universal tendency, the phenomenon that she terms *analytic support*:

> "In cognitively more demanding environments which require an increased processing load, language users – when faced with the option between a synthetic and analytic variant – tend to compensate for the additional effort by resorting to the analytic form." (Mondorf 2003: 253)

According to Mondorf (2003: 254), one of the effects of the so-called *more*-support that reflects the general *analytic support* is that "a separate lexeme as degree marker rather than an inflectional suffix can serve both as an unambiguous signal indicating increased processing load to the reader and as a less condensed and more explicit way of structuring a complex phrase". Since the synthetic variant in *-er* allows recognition only after the adjective and its inflection have been processed, complex environments should call for early recognition and hence the analytic variant with *more* would be used in English for the comparative construction (Mondorf 2003: 255). However, differently from the English comparative construction where the comparative marker in the synthetic alternative follows the adjective and precedes the adjective in the analytic alternative, the locative marker in the Estonian locative constructions follows the noun in both cases. That is, both the adessive case marker *–l* as well as the adposition *peal* come at the end of the noun phrase. Consequently, the arguments put forward by Mondorf (2003) do not directly work for this specific synthetic-analytic variation phenomenon in Estonian.

Given that the term weight can refer to either length or complexity, there are different ways **how weight can be measured**. Most studies count the number of words (Szmrecsanyi 2004, Rosenbach 2005: 632, et al. 2007, Hinrichs and Szmrecsanyi 2007, Bresnan Shih et al. to appear), but in other studies weight is defined in terms of number of syllables (e.g. McDonald et al. 1993), character counts (Wolk et al. to appear: 11) or in terms of phonological complexity (e.g. the number of lexical stresses as in Anttila et al. 2010). It is important to note that the different operationalizations are not identical. Taking the examples *laual* 'on the table' and *kirjutuslaual* 'on the desk', we can see that the two Landmark phrases do not differ in the number of words (in both cases only one word), but they clearly do differ in the number of syllables (*laual*: two syllables vs. *kirjutuslaual*: five syllables). Sometimes in the literature a distinction is also made between syntactic complexity (counting the number of words) and phonological complexity (counting the number of syllables or lexical stresses). I agree with Rosenbach (2005: 632) that it would be interesting to test explicitly for the effect of syntactic complexity versus length by comparing, for example, cases that differ in the syntactic complexity (i.e. the number of words) but not in the number of syllables/characters and vice versa.

In a more recent study, Wolk et al. (to appear: 11) report that several operationalizations of length were explored in their study on the English genitive alternation – the number of words, the number of characters and aggregated measures such as length differences and ratios between the possessor and the possessed constituents. Wolk et al. (to appear: 12) also coded a smaller data sample for the number of syllables in order to compare different operationalizations. Wolk et al. (to appear: 12) report that the correlation between syllable and word counts was 0.987 and between syllables and characters 0.993. For their analysis they opted to measure the length of both constituents in orthographic character counts. Hinrichs and Szmrecsanyi (2007: 453 Ftn.23) stress that they use length of a phrase in words as a "proxy for weight" because this is a method that has tradition in the study of weight effects in genitive choice and that it "strikes [them] as rather unproblematic". It is, however, important to bear in mind that English is a typologically different language than Estonian and what might turn out to be the most effective and best (or simply unproblematic) measure for length in English may not be so for Estonian.

Since we saw above that there are conflicting results as pertains to what is considered the best measure of length and since depending on how you count the length of a constituent, a distinction can be made between syntactic complexity and phonological complexity, it was thought best to use both measures in the present thesis. The variable 'length of the Landmark phrase' is thus measured in both syllables (phonological complexity) and words (syntactic complexity). An important methodological question concerns whether to include the monosyllabic adposition *peal* in the count as a separate word or not. Hinrichs and Szmrecsanyi (2007: 453) and Rosenbach (2005: 623) indicate in their studies on the English genitive alternation that the definite or indefinite articles determining the possessed phrase of an *of*-genitive were not included because it

provides a natural imbalance and skews the results. For similar reasons, it was decided not to include the adposition *peal* in the counts. Hence, Landmark phrases such as *laual* [laud+ADE] and *laua peal* [laud+GEN peal] were both considered as one word long and disyllabic. Following Wolk et al. (to appear: 12) and Bresnan and Ford (2010: 9 Ftn. 8) a logarithmic transformation is applied to the counts of words and syllables in order to compress extreme values and reduce skewness. Taking (31) as an example, the number of words in the Landmark phrase is 2 (*kušeti ääre* 'the edge of the couch' and the number of syllables is 5 (*ku-še-ti-ää-re*); after logarithmic transformation, the length scores are 0.69 and 1.61 respectively.

(31) *Ruudi* *istub* ***kušeti*** ***ääre*** ***peal*** [...]
Ruudi.SG.NOM sit-PRS.3SG couch.SG.GEN edge.SG.GEN on
<BCE: fiction>
'Ruudi is sitting **on the edge of the couch** [...]'

Based on the above discussion on the principle of end-weight and Rohdenburg's complexity principle, both of which have been found to play a significant role in the English genitive, dative and particle placement alternations, no directly comparable predictions can be made about the Estonian locative alternation. The main reason is that unlike the English alternation phenomena, the Estonian alternation does not involve word order change. However, Mondorf's analytic support (2003: 253; cf. above) is expressed in such terms that it can be, in principle, extended to alternation phenomena that do not necessarily involve a change in word order sequence, i.e. the alternation between the adessive and the *peal*-construction. Following Mondorf (2003: 253) and Rohdenburg (2003: 205) we can thus predict that the analytic adpositional construction will be used in cognitively more demanding environments.

It is at this point where the operationalization of 'complexity' becomes crucial. In this dissertation I have taken length as proxy for complexity, but this leads to another line of argumentation, namely one based on the principle of economy (Haiman 1983) and Zipf's 'principle of least effort' (Zipf 1935). Haiman (1983) discusses the iconic and economic motivation in language from a typological perspective. It is possible to relate the Estonian analytic adpositional construction and the synthetic case construction to Haiman's notion of distance (1983: 781–782): if linguistic distance is defined simply as the number of syllables between them, then the distance between the Landmark and the locative construction is least when they are bound morphemes (i.e. the adessive case) and greater when they are separate words (i.e. the adpositional construction). Haiman (1983) claims that in a number of cases the formal distinction between expressions is iconically motivated in that formal distance corresponds to conceptual distance. For example, the formal opposition between transparent vs. opaque corresponds to the pragmatic opposition between unusual vs. familiar (Haiman 1983: 802). The motivation for reduction and opacity is presumably economic and the notion of conceptual distance is related

to Zipf's (1935) principle of least effort. According to the Zipfian principle the more complex (in my data the adpositional construction) should be used less frequently than the less complex (the adessive construction). In other words, Zipf (1935) and Haiman (1983) provide further support for the general claims made in literature (Comrie 1986, Luraghi 1991, Hagège 2010, Lestrade 2010) that the case construction is more abstract, less specific and expresses more frequent relations than the adpositional construction.

At the same time, Haiman (1983: 812–815) also discusses the problem of competing motivation, i.e. instances where the iconic and economic motivations conflict with each other. Haiman (1983: 814) points out that "the observation that economy contrasts and conflicts with iconicity is not original" and that such observations have been made by "countless other scholars". For example, Ladefoged (1982: 241–242) makes a contrast between the principles of 'ease of articulation' and 'sufficient perceptual separation'. Transferring this to the alternation between Estonian synthetic locative constructions and analytic adpositional constructions, it can be claimed that the adessive takes the least effort and is easier to produce than the adpositional construction, but the adpositional construction allows for extra perceptual separation and more explicit expression of location.

Given the above discussion on iconicity and economy[22] and the complexity principle, it is possible to propose two sets of predictions. First of fall, it may be predicted that for reasons of language economy language users will prefer the more compact adessive construction in longer environments. At the same time, it is possible that for reasons of iconicity and the complexity principle the more explicit adpositional construction is preferred in longer environments. Length is here taken as a proxy for explicitness, i.e. longer Landmark phrases are taken to describe discourse new entities that are opaque and/or less familiar. The two possible explanations are in conflict, but it becomes clear from the results presented in Chapter 4 that the principle of language economy wins over the principle of analytic support favouring the explicitness or iconicity explanation in the alternation between adessive and adposition *peal* 'on' – the adessive case construction is preferred with longer and more complex Landmark phrases. At the same time, the experimental results lead to a slightly different conclusion (see Chapter 5 and the results of the sentence rating task).

Morphological complexity of Landmark. In addition to looking at the length of the Landmark phrase (i.e. quantity), the morphological composition of the phrase (i.e. quality) was likewise included. For each Landmark in the dataset it is established whether it is a 'simple lexeme' as *laud* in example 32, or a 'compound' as *kirjutuslaud* 'writing desk' in example 33. Similarly to the above argumentation and discussion about length of the Landmark phrase, morphological complexity of the Landmark phrase is taken as a proxy for

[22] See also the discussion in Rosenbach (2003: 399–402) on the iconic vs. economical tendencies in language and their relation to the English genitive alternation.

general complexity and based on Mondorf's (2003: 205) analytic support, it is predicted that the analytic, i.e. more explicit, construction is used in cognitively more demanding environments.

(32) ***Laua*** *peal* *aga* *seisis* *mitut* *arvutit*
table.SG.GEN on but stand-PRS.3SG many.SG.PRT computer.SG.PRT
meenutav *aparaat.* <BCE: fiction>
remind-PRS.PCPL machine.SG.NOM
'But a machine resembling many computers was standing **on the table**.'

(33) *Trengi* *sõrmed* *alustavad* *tantsu* ***kirjutuslaual.***
Treng.SG.GEN finger.PL.NOM begin-PRS.3SG dance.SG.PRT writing.desk.SG.ADE
<MDCE: fiction>
'Treng's fingers begin a dance **on the writing desk**.'

Syntactic function of the Landmark phrase. Both the adessive and the adpositional construction can fulfil two syntactic functions in a clause – that of an adverbial, as *õue peal* 'in the yard' in example 34, or a modifier as *merel* 'on the sea' in example 35. It is predicted that the adessive case will be the preferred construction when the locative phrase functions as a modifier for various reasons. Estonian adverbial modifiers follow the head word and it is claimed in Estonian reference grammars that postmodifiers are not very common in Estonian and they make the clause clumsy (Erelt et al. 2007: 535). The construction with the adposition is longer and has greater syntactic "weight", i.e. it is potentially even clumsier than the case construction in postmodification and should thus be avoided. The variable has two levels: 'adverbial' and 'modifier'.

(34) *Kuule,* *sul* *kasvab* ***õue*** ***peal*** *üks*
listen-IMP you.ADE grow-PRS.3SG yard.SG.GEN on one.SG.NOM
suur *puu?* <BCE: fiction>
big.SG.NOM tree.SG.NOM
'Listen, you have a big tree growing **in the yard**, don't you?'

(35) [...] *kaluritele* *ja* *teistele* ***merel*** *töötavatele*
fisherman.PL.ALL and other.PL.ALL sea.SG.ADE working.PL.ALL
inimestele [...] <MDCE: fiction>
person.PL.ALL
'[...] to fishermen and other people working **on the sea** [...]'

Word class of Landmark. Different expression types have been found to affect the choice of syntactic alternatives; see, for example, Bresnan and Ford (2010) for an overview of how this variable affects the dative alternation in English and Gries (1999) for the English particle placement. Since pronouns tend to be shorter words than nouns and they are less specific than full noun phrases, it may be plausible that the adpositional rather than the case construction is the

preferred construction with pronouns as Landmarks. This prediction is based on the assumption that the adpositional construction, by introducing an extra word *peal*, makes it easier for the speaker/writer to express the function intended. This variable has two levels – 'noun' (e.g. *öökapp* 'night stand' in example 36) and 'pronoun' (e.g. *see* 'this' in example 37).

(36) *Ka* *pruudikimp* *oli* *reaalselt* *olemas,*
also bridal bouquet.SG.NOM be-PST.3SG realistically be-PRS.SUP
seisis *käeulatuses* ***öökapil***. <MDCE: fiction>
stand-PRS.3SG within reach night.stand.SG.ADE
'The bridal bouquet was also realistically there, it was standing within reach on the **night stand**.'

(37) *Pärast* *tehti* *nari* *seina* *küljest* *lahti* *ning*
later do-PST.PASS bunkbed.SG.NOM wall.SG.GEN on open and
avanes *võimalus* ***selle*** *peal* *istuda.*
open up-PRS.3SG opportunity.SG.NOM this.SG.GEN on sit-INF
<BCE: fiction>
'Later the bunk bed was opened from the wall and the opportunity of sitting on **it** opened up.'

Word class of Trajector. The word class of the Trajector had three levels: 'noun' (e.g. *pihlakad* 'rowans' in example 38), 'pronoun' (e.g. *midagi* 'something' in example 39), and 'verb' (e.g. *vaatasin* 'watched' in example 40).

(38) ***Pihlakad*** *kasvasid* *reas* *sellelsamal*
rowan.PL.NOM grow-PST.3PL row.SG.INE same.SG.ADE
põllupeenral. <MDCE: fiction>
field.SG.ADE
'**Rowans** were growing in a row on this very same field.'

(39) *Tal* *on* *kindlasti* ***midagi*** *hinge* *peal*
he.ADE be-PRS.3SG definitely something soul.SG.GEN on
[...] <BCE: newspapers>
'He definitely has **something** on his soul [...]'

(40) *Õhtul* *hiljem* ***vaatasin*** *diivani* *peal*
evening.SG.ADE later watch-PST.1SG couch.SG.GEN on
telekat [...] <BCE: newspapers>
TV.SG.PRT
'Later in the evening I **watched** TV on the couch [...]'

Case form of Trajector. Another formal aspect of the Trajector phrase that may play a role in the choice between the adessive and *peal*-construction is the case form of the Trajector. As we saw from above, there are 14 cases in Estonian, and in theory, the nominal and pronominal Trajector phrase can occur in any of them. Due to data sparseness, this variable does not have 14 levels; instead, it has the following four: 'nominative', 'partitive', 'other cases' and 'not applicable'. It was possible to code the case form of only those Trajectors that were either nominal or pronominal; otherwise the variable received the level of 'not applicable', as example 44 where the Trajector is a verb phrase – *oli kõige toredam* 'was the nicest'. The category 'other cases' refers to instances when the pronominal or nominal Trajector is in another case besides the nominative (*ränne* 'migration' in example 41) or the partitive (*varblasi* 'sparrows' in example 42). For instance, in example 43 the Trajector is in the allative – *kellele* 'to whom'.

(41) *Merel* *algab* *kauride* ***ränne***. <MDCE: newspapers>
sea.SG.ADE begin-PRS.3SG diver.PL.GEN migration.SG.**NOM**
'**The migration** of the divers begins at sea.'

(42) [...] *isegi* *laudadel* *sirtsub* ***varblasi*** [...] <MDCE: fiction>
even table.PL.ADE chirp-PRS.3PL sparrow.PL.**PRT**
'[...] even on tables, there are **sparrows** chirping [...]'

(43) *... nägi* *ta* *Dandot ,* ***kellele*** *oli*
see-PST.3SG he.NOM Dando.PRT who.SG.**ALL** be-PST.3SG
trepi *peal* *kuul* *pähe* *tulistatud.*
stairs.SG.GEN on bullet.SG.NOM head.SG.ILL shoot-PST.PTCP
<BCE: newspapers>
'... he saw Danto, **who** had a bullet shot into his head on the stairs.'

(44) *Kesal* ***oli*** ***kõige*** ***toredam*** *karjas*
fallow.SG.ADE be-PST.3SG most nice herd.SG.INE
käia. <MDCE: fiction>
go-INF
'To herd cattle **was the nicest** on the fallow.'

Clause type. It may be that the type of clause plays a role in the Estonian locative alternation. This variable is coded in the corpus dataset as 'main' (example 45) or 'subordinate' (example 46), depending on in which type of clause the adessive or *peal*-construction occurs.

(45) ***Vana*** ***sepikoja*** ***varemel*** *õitsesid*
old.SG.GEN smithy.SG.GEN ruins.ADE blossom-PST.3PL
põdrakanepid *nagu* *lilla* *meri* [...] <MDCE: fiction>
rosebay.PL.NOM like purple.SG.NOM sea.SG.NOM
'The rosebays were blossoming like a purple sea **on the ruins of the old smithy** [...]'

(46) [...] *ununenud* *mälestus* *sellest* *ajast,* *kui*
forgotten.SG.NOM memento.SG.NOM this.SG.ELA time.SG.ELA when
nad *seda* *veel* *tupes* ***vöö*** ***peal***
they.NOM this.SG.PRT still sheath.SG.INE belt.SG.GEN on
kandsid. <BCE: fiction>
wear-PST.3PL
'[...] a forgotten memento from the time, when they still wore it in the sheath **on the belt**.'

Word order and the basic clause patterns. Although Estonian is typically considered a language with a relatively free word order (Lindström 2005: 10), any (functionally oriented) linguist will agree that nothing is ever completely "free". In Estonian, as in many other European languages, it is common to begin a clause with the information that is already known to the speaker/listener and to provide the new information at the end of a clause. Another variable that has been shown to affect word order in Estonian is the principle of end-weight (see Lindström 2005: 23-24 and the literature referred to in Lindström's study; cf. also the discussion under the variable 'Length of the Landmark phrase' above). Lindström (2005: 24) points out that it is difficult to say which variable is more important in determining Estonian word order – discourse-functional properties such as topic and focus, i.e. the information structure, or the principle of end-weight. To my knowledge, there is no detailed quantitative study that makes use of the elegant and sophisticated statistical methods on Estonian word order preferences and it would be interesting to see which variable 'wins'. Estonian will definitely provide interesting typological data to the on-going debate between the adherents of 'end-weight' (e.g. Hawkins 1994, 2004; Wasow 1997, 2002) and 'discourse' (e.g. Chafe 1994, Lambrecht 1994), but the scope of the present thesis does not allow me to elaborate on this issue. I only mention the word order variable in passim relative to the Estonian locative alternation.

According to Estonian reference grammars (e.g. Erelt 2003: 93–94) there are two basic syntactic patterns of clauses in Estonian: normal and inverted types of clauses. Table 7 lists these patterns: the first column identifies the type of clause, the second column specifies the order of constituents (S stands for subject, V for verb, A for and adverbial, and Obl for an oblique object) and the third column provides an example.

The basic word order in the normal clause is SVA (a in Table 7), where the subject (S) is morphologically unmarked, the verb (V) agrees with the subject in person and number, and the verb may be followed, for example, by a subject predicative and/or a locational adverb (A). In the "inverted clauses" it is not the subject that comes first in the clause, as in normal clauses, but an adverbial (A) or an oblique object (Obl) expressing location, time, possessor or experiencer (Erelt 2003: 93). Lindström (2005: 10) refers to these clauses as specific clause types in which the word order sequence is conventionalised in Estonian. Such conventionalised word order sequences make it easier for the listener/reader to understand the clause, since in these clauses information structure plays a role:

what is already known comes at the beginning and what is new, at the end of the clause (Lindström 2005: 10). Examples of such specific clause types include existential clauses (b in Table 7), possessive clauses (c in Table 7), state clauses (d in Table 7), and experiential clauses (e in Table 7).

Table 7. Basic clause patterns in Estonian (Erelt 2003: 93-95)

Clause pattern	Order	Example
a. NORMAL CLAUSE	S V A	*Jaan on toas.* Jaan.SG.NOM be-PRS.3SG room.SG.INE 'Jaan is in the room.'
b. EXISTENTIAL CLAUSE	A V S	*Aias on lilled.* garden.SG.INE be-PRS.3PL flower.PL.NOM 'There are flowers in the garden.' (lit. 'In the garden are flowers.')
c. POSSESSIVE CLAUSE	Obl V S	*Jaanil on vend.* Jaan.SG.ADE be-PRS.3SG brother.SG.NOM 'Jaan has a brother.' (lit. 'On Jaan is a brother.')
d. STATE-CLAUSE	A V S	*Väljas on kõva tuul.* outside be-PRS.3SG strong.SG.NOM wind.SG.NOM 'There is a strong wind outside.' (lit. 'Outside is a strong wind.')
e. EXPERIENTIAL CLAUSE	Obl V S	*Jaanil on hirm.* Jaan.SG.ADE be-PRS.3SG fear.SG.NOM 'Jaan is afraid.' (lit. 'On Jaan is fear.')

From the perspective of the present research question – the alternation between the adessive and the adposition *peal* – several points need to be kept in mind with regard to Estonian clause patterns. First of all, Estonian existential clauses (b), possessive clauses (c) and experiential clauses (e) have the same basic structure – the constituent expressing location (*aias* 'in the garden' in b), possessor or experiencer (*Jaanil* 'Jaan' in c and e) comes first. Secondly, while in the possessive and experiential clauses this initial constituent has to take on the adessive case, initial locative adverbials in the existential clauses can be expressed by the adessive or the adposition *peal* or any other locative construction.

Word order is considered in respect with the following two variables – position of the Landmark phrase within the clause and the relative position between the Landmark and Trajector. I describe these two word order variables below.

Position of the Landmark phrase. From the perspective of the alternation between the adessive case and the adposition *peal* it is important to note that in

principle, both constructions can come either at the beginning of a clause (examples 47a and 47b), in the middle of a clause (examples 48a and 48b), or at the end of a clause (examples 49a and 49b).

(47) INITIAL POSITION:

a. ***Puutrepil*** *istus* *peremees* <MDCE: fiction>
wooden.stairs.SG.ADE sit-PST.3SG landlord.SG.NOM
'The landlord was sitting on the wooden stairs.' (Lit. '**On the wooden stairs** was sitting the landlord.')

b. ***Trepi*** ***peal*** *istus* *Aime* *otsekui*
stairs.SG.ADE on sit-PST.3SG Aime.NOM like
väike *vanaeit* [...] <BCE: fiction>
small.SG.NOM old.woman.SG.NOM
'Aime was sitting on the stairs like a small, old woman [...]' (Lit. '**On the stairs** was sitting Aime like a small, old woman [...]')

(48) MIDDLE POSITION:

a. [...] *tundsin* *kuidas* *mu* ***laual*** *lebav* *käsi*
feel-PST.1SG how I.GEN table.SG.ADE lie-PRS.PTCP hand.SG.NOM
värisema *hakkas* <MDCE: fiction>
shake-INF start-PST.3SG
'[...] I felt how my hand, which was lying **on the table**, started to shake.'

b. *Lehevirnu* ***laua*** ***peal*** *oli* *mitu* [...]
paper.pile.PL.PRT table.SG.GEN on be-PST.3PL several
<BCE: fiction>
'There were several piles of paper **on the table** [...]'

(49) FINAL POSITION:

a. *Liikumatu* *muie* *püsis* ***ta*** ***näol***.
immovable.SG.NOM smirk.SG.NOM stay-PST.3SG he.GEN face.SG.ADE
<MDCE: fiction>
'An immovable smirk stayed **on his face**.'

b. *Mul* *oli* *mingi* *muhk* ***näo*** ***peal***. <BCE: fiction>
I.ADE be-PST.3SG some bump.SG.NOM face.SG.GEN on
'I had some kind of bump **on my face**.'

Based on the principle of end-weight which states that "long, complex phrases tend to come at the ends of clauses" (Wasow 1997: 81), I assume that it is the analytic adpositional construction with *peal* 'on' that creates a heavier constituent because it has the extra lexeme (*peal* 'on') and that the adpositional construction should thus prefer the clause-final position. This prediction finds further support in information structure – given that the adpositional construction is more explicit and specific than the case construction, it should be used at the end of clauses where new information is provided in Estonian. The case construction is predicted to be used at the beginning of a clause,

because it is shorter than the adpositional construction and less specific. Extra motivation for predicting the clause-initial position of the adessive constructions comes from Table 7 where it can be seen that this construction is used clause-initially in other constructions as well. Hence, the general frequency of the adessive construction to be placed in the initial position may play a role in the locative alternation as well.

Relative position of the Trajector and Landmark phrase. There is another aspect that should be pointed out in regard to word order. As was stressed above in the section on semantic variables, Tajector is considered to be the most prominent participant in locative expressions (Langacker 2008: 70). It is in the Landmark rather than in the Trajector where new information is found. It stands to reason, therefore, that the preferred word order for both constructions is such that the Trajector phrase precedes the Landmark phrase. Clauses were coded for the relative order of Trajector and Landmark phrases. The Landmark phrase either followed (example 50, coded as 'tr_lm')[23] or preceded (example 51, coded as 'lm_tr') the Trajector phrase.

(50) *Maja* *asus* ***kõrge*** ***mäekalda***
house.SG.NOM be.located-PST.3SG high.SG.GEN mountain.slope.SG.GEN
peal. <BCE: newspapers>
on
'The house was located **on a high mountain slope**.'

(51) ***Põrandal*** *lebab* *paks* *meeste* *Vogue*
floor.SG.ADE lie-PRS.3SG thick.SG.NOM man.PL.GEN Vogue.SG.NOM
<MDCE: newspapers>
'A thick men's Vogue is lying on the floor.' (Lit. '**On the floor** is lying a thick men's Vogue.')

Although the two operationalizations of word order are to some extent correlated, there is no complete one-to-one correspondence between the two variables. The variable 'position of the Landmark phrase' takes into account only the position of the Landmark within the clause irrespective of where the Trajector is positioned, while the variable 'relative position of Trajector and Landmark' specifies the position of the Landmark relative to the Trajector irrespective of where the phrase itself is positioned within the clause. For instance, the *peal*-construction (*ratta peal* 'on the wheel') in example 52 receives the coding 'final' for the variable 'position of the Landmark phrase' and 'lm_tr' for the variable 'relative position of Trajector and Landmark'.

[23] The Landmark phrase is given in bold; the Trajector phrase is underlined.

(52) *... otsekui* ***ratta*** ***peal,*** *kus* *ta* *tegelikult*
as.if wheel.SG.GEN on where he.SG.NOM actually
oligi ... <BCE: fiction>
be-PST.3SG
'... as if on a wheel where he actually was ...'

2.3.3 Variables not included

This section discusses some of the variables that are not included in the present analysis, but which are reported to play a role in other variation studies, and gives reasons for excluding them from this dissertation. It is hoped that future research takes these potentially relevant variables also into account. The variables that are left out fall into the following broad categories: phonology, variables related to discourse and sociolinguistic variables.

In addition to these broader categories, a number of variables or features discussed in the literature on space have been left out as well. These include variables such as 'orientation' and '(degree of) contact'. Although admittedly these may play a role in the alternation between the Estonian adessive and *peal*-construction, they were left out from the present study for reasons of the nature of the data and focus of the study. The analysis of the data includes, to a large part, corpus data. As has been already mentioned earlier, it is notoriously difficult to assess spatial expressions vis-à-vis the different spatial features without an accompanying image (cf. also the variable 'relative size of the Trajector in relation to the Landmark'). In addition, the focus of the present study is on providing a general overview of the alternation. At the same time, this line of research is clearly desirable in the future, albeit with a different dataset (e.g. an elicitation study using pictures as stimuli with the different spatial features as explanatory variables).

Phonological variables. There are not many alternation studies that explicitly consider the effect of phonological variables, with the exception of Antilla et al. (2010) and Shih et al. (to appear). The studies conducted by Antilla et al. (2010) and Shih et al. (to appear) demonstrate that phonological variables play a role in the English genitive and dative alternations. The results of Antilla et al. (2010) suggest that prosody significantly affects the choice of dative construction in spoken and informal written English. Shih et al. (to appear) found that while **rhythm** significantly influences the genitive construction choice in spoken English, its explanatory role is small relative to other predictors. They conclude, therefore, that "rhythm – and phonological variables at large – must not be discounted in studies of syntactic variation, but the converse is also crucially true: rhythm alone does not do or explain everything" (Shih et al. to appear: 1).

In addition to the two studies I have come across that look at phonology explicitly, there are a number of studies that look at the influence of a phonological variable as one among many others. Szmrecsanyi (in press b: 12), for example, reports that a **final sibilant** in the possessor NP is claimed to

encourage usage of the *of*-genitive due to a haplology or *horror aequi* effect – speakers tend to avoid immediately adjacent sibilants. Similar findings about the effect of final sibilancy are reported in other studies on the English genitive alternation, e.g. Hinrichs and Szmrecsanyi (2007). **Stress** is another phonological variable that has been discussed in alternation studies. Gries (1999: 109) points out that with strongly contrastively stressed direct objects, it is the direct object that follows the particle in his study on particle placement in English. Mondorf (2003: 274–283) describes three instances in which phonology causes increased processing complexity, which is compensated by means of the analytic comparative construction in English. She reports that the analytic comparative can create an additional emphasis (Mondorf 2003: 294). This supports the claim made by Bolinger (1986: 58) who states that "[f]unction words may be contrastively accented, which is hardly possible with inflections". Given that *peal* is a function word and the adessive an inflection, it is possible for the analytic locative construction in Estonian to be contrastively accented and to be the more stressed alternative of the two.

The reasons for excluding prosodic and phonological variables from the corpus studies described here is that this area is largely unexplored in Estonian grammar (Ross et al. 2003). A further important reason for not studying the effect of these variables is that unlike the data samples described in Antilla et al. (2010) and Shih et al. (to appear), my data samples do not include spoken language, but are based on written Estonian (see Chapter 3 for the detailed description of the datasets). Shih et al. (to appear: 20) assume that rhythmic and phonological effects will be most apparent in spoken contexts and that the role of rhythm in spoken and written language may differ – speech is spontaneous, but writing is calculated. They (Shih et al. to appear: 20) also note that writers may not be as worried about phonological properties in written work and that the effect of phonological variables may be greater in spoken use.

Discourse-related variables. Other variables that are not considered in this study pertain to different aspects of the discourse and include the following: definiteness, topicality, discourse accessibility, lexical density, structural parallelism, and text type. **Definiteness** has been shown to play a role in the dative alternation (e.g. Bresnan and Ford 2010: 174-175). Since Estonian does not have a system of definite and indefinite articles and measuring the definiteness of either the Landmark or the Trajector phrase is hence not as straightforward as, for instance, in English, this variable was left out of the present study. However, the predictor 'word class' introduced above (cf. pronouns vs. nouns) addresses some of the similar questions and provides insight into whether definiteness might play a role in the alternation between locative cases and adposition in Estonian or not.

Hinrichs and Szmrecsanyi (2007: 451) present evidence that the typical possessor head noun of an *s*-genitive has a higher **text frequency** than the typical possessor head noun of an *of*-genitive. Studies on the English genitive (Hinrichs and Szmrecsanyi 2007, Szmrecsanyi and Hinrichs 2008, Szmrecsanyi

in press b) and adjectival comparative construction (Mondorf 2003, Szmrecsanyi 2005) show that the synthetic construction is preferred with frequently used nouns and adjectives. Mondorf (2003: 260) points out that frequently used adjectives are typically also the shorter ones and tend to favour the *–er* variant. Mondorf (2003) takes a general stand on complexity and ties all of the variables she discusses with her notion of analytic-support. She claims that the retrieval of well-entrenched lexemes from the lexicon or their on-line construction by means of adding *–er* takes less processing effort and that the relative difficulty experienced in accessing rare lexical items is compensated by using the analytical *more* construction (Mondorf 2005: 260). It may be predicted that frequently used Landmarks prefer the locative case construction while more infrequent Landmarks prefer the analytical *peal*-construction. Validating the significance of such frequency effects is left for future research, although the corpus studies address the issue of frequency in passing.

There is an abundance of research demonstrating that **discourse accessibility** influences the choice of alternative constructions (see, for example, studies cited in Bresnan and Ford 2010: 174). It is posited for the English word order alternations that one of the constructions is preferred due to the different information status the alternative constructions typically assign to the respective constituents. For instance, it has been shown that the *s*-genitive is preferred if the possessor is given because it places the given element first (Hinrichs and Szmrecsanyi 2007: 451). Shih et al. (to appear: 14), however, report a conflicting result – they find that there is a tendency for given possessors to occur in the *of*-genitive construction. Although there is a reversal of expectations, this variable is still significant as a whole. Gries (1999: 111) reports that topicality or givenness of the direct object explains why pronouns and referentially vague (or empty) nouns require that the particle is placed after the object, whereas heavily modified nouns most frequently occur in the VPO construction.

Lexical density is another discourse-related variable that has an effect on various alternations. Szmrecsanyi (2006), Hinrichs and Szmrecsanyi (2007), Szmrecsanyi and Hinrichs (2008) have shown that speakers prefer the *s*-genitive in contexts characterized by high type-token ratios, which have taken to be indicative of increased lexical density. Hinrichs and Szmrecsanyi (2007: 457) provide evidence that when news writers feel a need to economically code more information in a textual passage, they use the *s*-genitive since it is more compact and economic than the analytic *of*-genitive. Hinrichs and Szmrecsanyi (2007: 457) operationalized lexical density by establishing the type-token ratios of the textual passage where the genitive occurrence was embedded. Their cut-off point was 50 words before and 50 words after the construction. Similarly to Szmrecsanyi (2006), they consider type-token ratio a proxy for lexical density: the more different word types are present in a given passage, the higher the lexical density and the more pressing the need to code economically, i.e. use the *s*-genitive.

Szmrecsanyi (2005, 2006), Hinrichs and Szmrecsanyi (2007), Bresnan and Ford (2010: 174) show that **structural parallelism** or **persistence** is an important variable in syntactic choice. According to Szmrecsanyi (2005: 113), speakers re-use a recently used or heard linguistic construction whenever they can and factoring in this predictor increases the researcher's ability to account for linguistic variation. This phenomenon has been called 'persistence' (e.g. Szmrecsanyi 2005, 2006; Hinrichs and Szmrecsanyi 2007), 'priming' (e.g. Bock 1986), 'structural parallelism' (e.g. Bresnan and Ford 2010), or simply 'repetition in discourse' (e.g. Tannen 1989). Research on the English genitive alternation (Szmrecsanyi 2006: 87-101; Hinrichs and Szmrecsanyi 2007) has shown that this kind of priming effect is present in both spoken and written English, although Shih et al. (to appear: 10) report that in their spoken English dataset the measure of persistence was an insignificant variable in predicting genitive choice. Gries (2005) provides further evidence that syntactic priming, i.e. the tendency to reuse syntactic constructions, plays a significant role in the English dative alternation and particle placement of transitive phrasal verbs.

In addition, **formality** or text type has been reported to determine the choice between the *s*-genitive and *of*-genitive in English with the *s*-genitive being more frequent in informal text types (Dąbrowska 1998, Rosenbach 2003, Hinrichs and Szmrecsanyi 2007). Another argument put forward in the English genitive literature is that because the *of*-genitive is of Anglo-French origin, it is deemed more formal than the more informal Germanic *s*-genitive (Shih et al. to appear: 11). No conclusions can be drawn about the Estonian locative alternation due to the procedure used to select the data. In order to determine if and how the Estonian locative alternation differs across registers, a different approach is necessary to arrive at a more representative data sample.

The discourse-related variables described above were excluded from the present study for various reasons. The aim of the corpus and experimental studies included in the present thesis is to paint a general picture of the alternation phenomenon in written Estonian by placing emphasis more on the various linguistic variables, both semantic and formal, than on extralinguistic variables like discourse and register.

Sociolinguistic and diachronic variables. Hinrichs and Szmrecsanyi (2007), Szmrecsanyi (in press a), Wolk et al. (to appear) have shown that language history and language variety also play a crucial role in different alternation phenomena. In addition to describing the variables that determine construction choice in present-day English, Szmrecsanyi (in press a) and Wolk et al. (to appear) demonstrate how **diachronic data** can be valuable in studying alternation phenomena. Such studies not only give a diachronic perspective on how the quality and quantity of the alternation itself has changed, but the combination of diachronic data and logistic regression analysis allow the researcher to assess if the effect of the variables has changed over time or not. Although we do not have a similarly large historical corpus available for Estonian as for

English, the study of Estonian locative alternation phenomenon from the diachronic perspective is possible.

Another sociolinguistic factor that plays a role at least in alternation phenomena in English is the language **variety** – the *s*-genitive is, for example, used more frequently in American than in British English (Rosenbach 2003: 384). In the context of Estonian, dialectal differences may play a role in the choice between a synthetic or analytic construction. Palmeos (1985: 15) points out that the use of analytical constructions is characteristic of Southern and Western Estonia. It will certainly be interesting to see if this claim can be verified by either experimental or corpus data and how the alternation between locative constructions relates to other phenomena that can be expressed either by a synthetic or an analytic expression.[24] Another "lectal" variable that I predict plays a significant role is idiolect. I discuss this issue in more detail in Chapter 5 where I present the results of the two rating tasks.

However, a more detailed study of variety and language history is left for future research. Establishing first the phenomenon in standard written Estonian is considered as a necessary stepping stone from which future research can spring forth taking into account the different lectal aspects discussed in cognitive sociolinguistics (Geeraerts *et al.* 2010).

2.4 Objectives and general predictions

Taking the previous studies on the alternation between cases and adpositions and other grammatical alternations described in this chapter into account, the aim of the present thesis is to provide a multivariate account on the alternation between the two locative constructions of Estonian in present-day written Estonian. Table 8 lists the semantic and morphosyntactic variables together with their levels studied in the dissertation.

Table 8. List of semantic and morphosyntactic variables examined in the thesis

Semantic variables	Levels
type of relation btw LM & TR	abstract, spatial
type of LM	place, thing
type of TR	abstract, object
animacy of LM	animate, inanimate
animacy of TR	animate, inanimate
number of LM	plural, singular
number of TR	plural, singular
mobility of LM	mobile, static
mobility of TR	mobile, static
relative sizes of TR & LM	conventional, same, unconventional
verb	action, existence, motion, posture, no verb

[24] See, for example, Uiboaed 2010 on phrasal verbs in Estonian dialects.

Morphosyntactic variables	Levels
length of LM phrase	syllables and words
morphol. complexity of LM	simple, compound
syntactic function of LM	adverbial, modifier
word class of LM	noun, pronoun
word class of TR	noun, pronoun, verb phrase
case form of TR	nominative, partitive, other, not applicable
clause type	main, subordinate
position of LM phrase	initial, middle, final
relative position of TR & LM	lm_tr, tr_lm

The first two studies described in the dissertation are corpus studies, while the second two are experimental studies that test some of the hypotheses posited on the basis of corpus studies and my own intuition as a native speaker of Estonian. The first corpus study is multivariate in nature and the aim is to identify the possible variables that determine the choice between the adessive case construction and the *peal*-construction. Both univariate and multivariate statistical techniques are used to analyse the results. The univariate analysis is seen as a necessary and beneficial first step in the quantitative analysis of the results. The purpose is to move from more simple, univariate exploratory analysis to more complex but more powerful multivariate analysis. Univariate techniques make it possible to identify which variables are statistically significant, but by employing such a multivariate statistical technique as regression analysis (see section 3.4 for details), it is possible to determine the contribution of the different semantic and morphosyntactic variables to the alternation and calculate the relative strength of each individual variable. The second corpus study focuses purely on the semantic variable of type of Landmark and employs the methodology of distinctive collexeme analysis. This method is used in order to see if there are specific lexemes that distinguish between the two constructions. Should there be such distinctive lexemes the question then arises whether these lexemes form any semantic groups that in turn give insight into the semantics of the two constructions.

As seen above, both constructions fulfil functions where the two are not interchangeable, e.g. the adessive fulfils the function of the logical subject where the *peal*-construction is not a possible alternative. However, in the two corpus studies described in the dissertation, only those instances of the two constructions were included where the two were interchangeable. The exact nature of the datasets and the extraction procedure are described in detail in the next chapter (section 3.1). Thus, it should be noted that the aim is not to give an overview of the usage patterns of all of the instances of the two constructions, i.e. the reader should not expect a prototype or schematic analysis of the adessive case or the adposition *peal* 'on'. Such a general analysis of Estonian locative cases and adpositions is a desirable objective of future research.

The general prediction of the two corpus studies is that although both constructions are used to express support-relations in Estonian, i.e. spatial

relations where one object is placed on top of another object, there are certain semantic and morphosyntactic constraints that influence the alternation between the two. Since no previous large-scale quantitative study exists specifically on this alternation, the corpus studies are exploratory in nature rather than specifically hypothesis testing. This is where experimental linguistics enters the stage.

The two rating tasks included in the dissertation test the validity of three hypotheses that are proposed on the basis of the corpus data, the studies on other alternation phenomena, and my own intuition as a native speaker of Estonian. The first study, picture rating task, looks at the semantic variable type of Landmark and the morphosyntactic variable word order. Both corpus studies suggest that there is a difference between the case construction and the adpositional construction as to which types of Landmarks are used with these constructions. A similar conclusion is drawn by studies on the alternation between the synthetic and analytic locative constructions in the Saami language (Bartens 1978) and the Finnish language (Ojutkangas 2008). The synthetic case construction tends to be used with larger locations and the analytic adpositional construction with smaller locations. The second rating task looks at two morphosyntactic variables – length of the Landmark phrase and word order. The reason why both rating tasks include the variable word order has to do with converging evidence and testing the validity of native speaker intuitions. Although in the multivariate corpus study the word order variable does not play a particularly decisive role, it was decided to validate this result with experimental results. The results of the picture rating task indicate that the word order variable is a significant, although not a particularly strong, variable in the alternation between the two locative constructions.

The second rating task, however, points towards a different conclusion – there seems to be no specific interaction between the type of locative construction and word order. The variable 'length' is included in the sentence rating task because both the univariate and multivariate results of the first corpus study indicate that this is one of the strongest predictors in the alternation between the synthetic case construction and the analytic adpositional construction.

The next chapter (Chapter 3) outlines the exact designs of the studies, while Chapters 4 and 5 present the results for corpus and experimental studies respectively. Detailed discussion on both converging and diverging results is provided in Chapter 6.[25]

[25] The reason why this somewhat unusual organisation is chosen in presenting the methods and results of the four different studies is to avoid repetition. The studies share, for example, the statistical techniques and the general design. This makes it more convenient to discuss all the issues related to methodology together in one chapter and to leave the reporting of the results for the two remaining chapters, based on the nature of data (corpus vs. experimental studies).

3. RESEARCH METHODS

The previous chapter describes in detail the linguistic phenomenon under study, outlining the response or dependent variable (the Estonian adessive case and the adposition *peal* 'on') as well as the numerous explanatory or independent variables that are predicted to explain the alternation (the different semantic and morphosyntactic variables and their values listed in section 2.3). The next step in a scientific investigation is data collection and analysis. However, before the actual results can be presented, it is necessary to describe the specific methodology employed to collect the data and the statistical techniques used in the subsequent analysis stage. In general, the aim of data analysis is statistical inference, i.e. to arrive at a model of the population[26] based on the data sample using the induction process (Cohen 1988: 10). There are two approaches to statistical inference: a researcher can either estimate the values of the population parameters in a model from the data or posit specific hypotheses about the values of the population parameters and examine whether the data are consistent with the hypotheses or not. In the case of **estimation**, the aim is to describe the variability in the estimates population based on the probability estimations arrived at evaluating a specific sample. This is the general goal of the corpus studies. The aim of the second approach, **hypothesis testing**, is to adopt specific values for population parameters, i.e. posit specific hypotheses which are to be accepted or rejected. This is the general goal of the two linguistic experiments. Every data sample, of course, gives a slightly different picture of the population and statistical techniques are needed to assess the degree of chance in the results.

The present chapter starts with the description of the corpus studies and the experimental designs. The second part of the chapter gives a short overview of the univariate and multivariate[27] statistical techniques employed to analyse the results of corpus and experimental studies. Section 3.1 outlines the two corpus studies – the first is referred to as the multivariate corpus study[28] (section 3.1.1) and the second as the distinctive collexeme analysis (section 3.1.2). Both of these sections describe the exact procedure for data extraction as well as the general characteristics of the specific methodologies. The aim of both corpus studies is to establish the general usage patterns of the two constructions in written present-day Estonian. The first study focuses on a larger set of different semantic and morphosyntactic variables that are predicted to play a role in the alternation between the two constructions. The second study, however, looks at

[26] In the present case, the population under study is the alternation between the adessive case and the adposition *peal* 'on' in present-day written Estonian.

[27] Following Arppe (2008) I have used the terms *univariate* and *multivariate* instead of the terms *monofactorial* and *multifactorial*. The latter terminology is used by Gries (2009), but as Arppe (2008: 71 Ftn. 36) points out, these may cause confusion with *Factorial Analysis* as a statistical method.

[28] The multivariate corpus study described here supersedes the earlier version of the study published as Klavan (to appear).

a larger data set and takes a purely semantic point of view. It examines, among other things, the issue of frequency. The aim is to find out which lexical elements are typical for the two constructions and this collocational information is taken as evidence for the semantics of the constructions. Section 3.2 describes the designs of the two experimental studies. The section starts off with a general discussion of the reasons why these specific designs were used and highlights some of the advantages and disadvantages of using rating tasks. A selection of univariate and multivariate statistical techniques are used to analyse the corpus and experimental results; these are described in sections 3.3 and 3.4 respectively.

3.1 Corpus studies

Since the aim of the thesis is to present a list of possible methodological tools that can be employed to study variation phenomena like the one described here – the alternation between locative cases and adpositions –, it is not the size of one specific study (cf. for example, the 8,300 instances of genitives studied by Hinrichs and Szmrecsanyi 2007)[29] that matters so much, but the fact that more than one methodology should be used in order to assess how well it is suited for the context of alternation phenomena in Estonian and what the nature of the results obtained via different methodologies is. Therefore, the corpus studies described in this section do not include thousands and thousands of occurrences of the two constructions (except for the distinctive collexeme analysis), but rather serve as the foundation on which other larger corpus studies can be built. Another aim of the thesis is to provide a sketch of the two constructions in present-day written Estonian; hence, the choice of data in the corpus studies is written language, more specifically fiction and newspaper texts.

3.1.1 Corpus study I: Multivariate Corpus Analysis

A multivariate corpus analysis takes into account the multifaceted nature of everyday language use by considering all of the studied variables or factors simultaneously. Corpus occurrences of the two constructions were manually coded for different morphosyntactic and semantic variables (see Table 8 on p. 96) and entered into logistic regression analysis, which determines the relative influence of all variables and yields a predictive model. The aim is to model the choice between the adessive construction and the *peal*-construction as a function of a wide range of variables or predictors as they are referred to in regression terms. As mentioned above, this type of corpus analysis falls into the first type of statistical inference. Rather than testing the validity of previously

[29] The reader is here referred back to section 1.3, where it was stressed that more data does not necessarily lead to better results. With very large datasets, the risk of commiting a Type I error (finding a significant result when it is not there) is considerably larger. In datasets that comprise thousands of examples, even the smallest difference becomes statistically significant, while the effect size itself is very small.

posited hypotheses, the aim is to arrive at a model that adequately describes the variability in the population based on the probability estimates arrived at by evaluating the specific corpus sample. Since there are relatively few previous studies available on this topic, the approach taken is to annotate the corpus extractions for anything potentially relevant. It therefore bears considerable resemblance to corpus studies of the Behavioural Profile style (e.g. Gries and Divjak 2009, Gries 2010). At the same time, as described in Chapter 2, previous studies on different variation phenomena and spatial language served as an inspiration for generating the list of explanatory variables.

The data. The data analysed for this study comes from the corpus of present-day written Estonian – the fiction and newspaper sub-corpora of the Morphologically Disambiguated Corpus (2010; size 215,000 words) and the fiction and newspaper sub-corpora of the Balanced Corpus of Estonian (2008; size 10 million words). The reason why two different corpora are used has to do with practical considerations, more specifically with the question of how to extract the adessive construction from the corpus and the relative infrequency of the adposition *peal* 'on' vis-à-vis the adessive case. The *peal*-construction can be extracted from any corpus of Estonian language with relative ease – one just has to insert the string of characters for the adposition in the query box at, for example, http://www.cl.ut.ee/korpused/grammatikaliides/ and the output includes a list of all the sentences with the string *peal*. The output, of course, contains noise and will be described in detail below. However, the only 'fool-proof' way to extract the adessive construction from a corpus, which is formally marked with *-l*, is to use a morphologically tagged corpus. At the current stage (as of 18 January 2012), the Estonian tagged corpus (referred to as the Morphologically Disambiguated Corpus, available at http://www.cl.ut.ee/korpused/morfkorpus/) is not a very large corpus – it includes 500,000 words from six different text types (104,000 words from fiction; 75,500 words from translated fiction; 111,000 words from newspapers; 121,000 words of legal texts; 98,000 words of scientific texts; and 4,000 words of reference texts). Since the morphologically tagged corpus did not provide enough instances of the adpositional construction with *peal* 'on', a larger corpus of standard present-day written Estonian was used – the Balanced Corpus of Estonian.

There is software available that can be used to tag any electronic text in Estonian (see, for example, http://www.filosoft.ee/html_morf_et/). Since the majority of the texts included in the various corpora of Estonian are available for download, then in principle, these texts could be downloaded to a PC and inserted for a morphological analysis through the above-mentioned web-page or using a Perl script. However, the results of such automatic tagging software still need heavy manual postediting – the general error rate of the output is around 3.5 per cent (Habicht et al. 2000: 624, Kaalep and Vaino 2000: 92). Although the overall error rate might not seem that high it is painfully relevant for the adessive case construction due to reasons of grammatical homonymy. Many

words in Estonian end with the letter *l* which, incidentally, is also the case ending for the adessive case. Hence, the automatic tagging software erroneously identifies words like *keel* 'language', *veel* 'still', *seal* 'there' and countless other words as the adessive forms of other words, like *kee* 'necklace' (adessive: *keel* 'on necklace'), *vesi* 'water' (adessive: *veel* 'on water'), *siga* 'pig' (adessive: *seal* 'on pig'). Another source of errors for the automatic tagging software are Estonian adverbs and adpositions. Many of them have the same ending as the adessive case and have, in fact, historically grammaticalized from the adessive forms of nouns (Habicht 2000, 2001a, 2001b; Ojutkangas 2001). Therefore, the automatic tagging software erroneously identifies adpositions and adverbs like *puhul* 'in case of', *kõrval* 'next to', *vahel* 'sometimes' and many others as the adessive case forms of nouns like *puhk* 'case', *kõrv* 'ear', *vahe* 'difference'. As we saw in Chapter 2 above, the word class of adverbs (and even adpositions) in Estonian is more an open than a closed class, and it is virtually impossible to improve the tagging software so that it would not make these mistakes. New adverbs are continuously created, often with the aid of the adessive case ending and sometimes, although infrequently, the two forms co-exist side by side with identical form – for instance, *varjul* 'in the shadow' as an adverb in example 53, and the adessive case form of the noun *vari* 'shadow' in example 54.

(53) *Aknad* *olid* *ilusti* *rehealuse* *nurgas*
window.PL.NOM be-PST.3PL nicely room.SG.GEN corner.SG.INE
varjul. <BCE: fiction>
shadow
'The windows were nicely in the corner of the room out of sight (lit. **on/in the shadow**).'

(54) ***Varjul*** *on* *mingi* *seos* *inimese*
shadow.SG.ADE be-PRS.3SG some connection.SG.NOM person.SG.GEN
hingega. < BCE: fiction >
soul.SG.COM
'The **shadow** has some kind of a connection with a person's soul.'

The output of the automatic tagging software thus includes a list of possible word forms and it is the task of the researcher to disambiguate between the different forms and select the right one for the specific context. This is also the reason why the Estonian morphologically tagged corpus is referred to as a 'disambiguated corpus' – it takes a lot of effort to disambiguate between the various instances of grammatical homonymy that abounds in the Estonian language and sometimes different researchers disagree on the correct tag (see e.g. Habicht et al. 2000). For these reasons, it seems that the effort required by the manual postediting of the output of such automatic tagging software almost equals the effort it would take to look for instances of the adessive case without any automatic corpus query, i.e. extracting the adessive constructions completely manually. Still, since for Estonian there is a morphologically tagged corpus available (small as it might be), it was decided to use this corpus for the

adessive construction; for the adpositional construction, however, a larger untagged corpus had to be used.[30] Since the two constructions come from two different corpora and the constructions were randomly selected, the two groups are treated as independent samples (see section 3.3 below for explanation) when different statistical methods are employed in the analysis of the results.

The following section describes the type of adessive and *peal*-constructions that were included in the corpus studies; it will also list the instances that were excluded from the studies. It is necessary to specify what in variationist literature has been termed the **variable context** (e.g. Hinrichs and Szmrecsanyi 2007: 444, Wolk et al. to appear: 4; cf. choice contexts in Rosenbach 2003: 383), since the inclusion or exclusion of certain instances may skew the data. It should be clear from the discussion in Chapter 2 that not every adessive case construction can be expressed by a *peal*-construction and vice versa. It is therefore crucial from the perspective of the dissertation to include only those contexts where the two constructions are truly variable or interchangeable before proceeding with any kind of quantitative analysis. However, it is worth pointing out that even within the variable contexts there can probably never be complete synonymy between the two constructions due to the principle of isomorphism as expressed by Haiman (1983) and Bolinger (1977).

Extraction procedure for multivariate corpus study. In order to collect valid data for the multivariate corpus study, the following procedure was undertaken. I first extracted all of the adessive case instances from the fiction and newspaper sub-corpora of the Morphologically Disambiguated Corpus of Estonian (a total of 4,168 sentences: 1,755 sentences from fiction and 2,413 from newspapers) and copied them to an Excel spreadsheet. I did the same with the *peal*-construction by extracting all of the instances of the string *peal* from the fiction and newspaper sub-corpora of the Balanced Corpus of Estonian (a total of 2,387 sentences: 1,874 from fiction and 513 from newspapers) and copied them to an Excel spreadsheet. Since the output of the corpus queries does not give a list of random hits, but arranges the sentences in an alphabetical order according to the reference code at the beginning of each corpus sentence, the random function available in Excel was used to generate a random list of all of the adessive and *peal*-constructions. The next step involved coding each and every occurrence of the adessive construction either 'interchangeable' or 'not interchangeable' with the *peal*-construction. In case there were two adessive or *peal*-constructions present in one and the same sentence, only the first occurrence was coded. I then repeated the procedure with *peal*-sentences. After this initial coding stage I ended up with 488 sentences that contained an instance of the adessive construction that was judged interchangeable (see the criteria below) with the *peal*-construction and a list of 1,308 sentences with an interchangeable *peal*-

[30] It is important to note here, once again, that the reason why the Morphologically Disambiguate Corpus of Estonian was not used as the source of *peal*-constructions as well is that there were not enough instances. The corpus query resulted in only 45 hits; however 450 occurrences were taken as the absolute minimum for statistical analysis.

construction. From these lists I selected a random sample of 450 adessive constructions and 450 *peal*-constructions. The selection of 450 instances of each construction for the corpus study is relatively arbitrary – the only criterion was that there should be enough instances of both constructions to run logistic regression analysis with the data and to manage with the manual tagging of the 20 explanatory variables. Table 9 gives the number of all adessive and *peal*-constructions that were extracted from the two corpora, as well as the number of tokens coded as interchangeable, the number of tokens coded as not-interchangeable and the number of tokens selected for the analysis (900 tokens in total). It also reports the size of both MDCE and BCE – the two sources for the adessive constructions and *peal*-constructions respectively.

Table 9. Raw frequencies of the two constructions vs. the number of tokens selected in multivariate corpus study

	Adessive construction	***Peal*-construction**
Total number of the construction in fiction sub-corpus	1,755	1,874
Number of tokens coded as interchangeable	330	1,139
Number of tokens coded as not-interchangeable	1,425	735
Size of the fiction sub-corpus	MDCE: 104,000 words	BCE: 5,000,000 words
Total number of the construction in newspaper sub-corpus	2,413	513
Number of tokens coded as interchangeable	158	169
Number of tokens coded as not-interchangeable	2,255	344
Size of the newspaper sub-corpus	MDCE: 111,000 words	BCE: 5,000,000 words
Total number of tokens in the two sub-corpora	**4,168**	**2,387**
Number of tokens selected for the analysis	**450**	**450**

It should be clear from both the discussion of the various functions the adessive case fulfils in the Estonian language (cf. section 2.1.1) and from Table 9 above that the adessive construction is much more frequent than the *peal*-construction – in order to extract 4,168 sentences with an adessive construction, one only

needs a corpus of 215,000 words, while in order to extract 2,387 sentences with a *peal*-construction, one needs a corpus as big as 10 million words. However, the number of adessive constructions that were coded as interchangeable with the *peal*-construction (330 out of 1,755, approx. 20 per cent) is considerably smaller than the number of *peal*-constructions coded as interchangeable with the adessive construction (1,139 out of 1,874, approx. 60 per cent). This asymmetry points to several things. First of all, the adessive frequently fulfils other functions in present-day written Estonian besides expressing space, while the adposition *peal* is predominantly used to do exactly that. Secondly, it highlights the relevance of polysemy in studying synonymy. It is plausible to predict that one of the main reasons why the adposition *peal* is used as an alternative for the adessive case is to avoid ambiguity with the other readings of the adessive. The locative function is more a core function for the less frequent *peal* than for the more frequent adessive, which may have implications for production and interpretation – *peal* may signal locativity more unambiguously than the adessive.

Another aspect that the figures in Table 9 point towards is that the *peal*-construction is considerably less frequent in the newspaper texts (160 interchangeable instances per 5 million words) than in fiction (1,139 interchangeable instances per 5 million words). One of the reasons may be the register differences between the two text types. Newspaper texts, referred to also as journalese, are a very particular register. According to Gilquin and Gries (2009: 7) such texts are created more consciously than other texts, they come with restrictions such as word length, they are often not written by a single person and they may be heavily edited, etc. It seems plausible to predict that due to these reasons (especially the restrictions pertaining to word and character length of newspaper articles) the adposition *peal* seems dispreferred in newspapers and is more frequent in fiction texts. The number of interchangeable adessive tokens is also considerably larger for fiction texts than for newspaper texts. This fact seems to indicate that newspaper articles talk less about spatial relations than fiction texts.

In the manual coding of the two locative constructions for mutual interchangeability I relied on my own native-speaker intuition[31]. Only these instances of the adessive case were retained which could have been expressed with the adposition *peal*. Similarly, only those *peal*-constructions were retained which could have been expressed using the adessive case. The most important criterion was that the alternative construction would leave the meaning of the actual choice unchanged. A negative list of noninterchangeable types and occurrences guided the judgements. In fact, in a great majority of instances, the **adessive construction** proved to be **noninterchangeable** with the *peal*-construction; the following cases were excluded from the analysis (the number in the brackets indicates the number of the specific type that is excluded):

[31] I am aware that such intuitions are subjective and that there are interspeaker differences in judging the grammaticality of alternations.

(i) constructions expressing a **logical subject** (1,451) – e.g. *Mul on kaks eurot* 'I have two Euros';

(ii) **temporal expressions** (1,502) – e.g. *talvel* 'in winter';

(iii) constructions expressing **manner** (158) – e.g. *tasasel häälel* 'in a soft voice';

(iv) **adverb-like noun forms** (373) – these adessive forms of certain nouns behave very much like adverbs; they are not, however, considered as 'true' adverbs in their own right and they do not have a separate entry in the Dictionary of Written Estonian (http://www.eki.ee/dict/ekss/). They are, therefore, identified as adessive case constructions in MDCE. The following forms were excluded from the analysis: *alusel* 'on the basis of' (29), *andmeil/andmetel* 'according to' (22), *arvamusel* 'in the opinion' (1), *arvel* 'on account' (1), *baasil* 'on the basis' (2), *eeldusel* 'on the assumption' (1), *hinnangul* 'in the opinion' (27), *juhtimisel* 'under the direction' (1), *kaalul* 'at stake' (1), *kaalutlustel* 'for reasons' (1), *kaugusel* 'away' (2), *kinnitusel* 'according to' (13), *kohal* 'in place' (10), *kõrgusel* 'at height' (1), *käsul* 'by somebody's order' (1), *loal* 'with permission' (1), *maal* 'in the country' (6), *määral* 'to an extent' (8), *nimel* 'in the name of' (1), *nõudel* 'by demand' (1), *palvel* 'by request' (1), *pool* 'at' (20), *põhjusel/põhjustel* 'for reasons of' (8), *rindel* 'on the front' (2), *selgitusel* 'on explanation' (1), *soovitusel* 'on recommendation' (1), *sõnul* 'according to' (149), *süül* 'by somebody's fault' (2), *tagajärjel* 'because of' (3), *tasandil* 'on the level of' (2), *tasemel* 'on the level of' (8), *taustal* 'against the background of' (4), *teatel* 'according to' (5), *teel* 'on the road (to)' (12), *tingimusel* 'on the condition' (1), *tulemusel* 'as a result of' (2), *vahendusel* 'according to' (2), *vajadusel* 'in case' (2), *väitel* 'according to' (6), *välismaal* 'abroad' (5), *õhutusel* 'by incitement' (1), and *õlakõrgusel* 'shoulder-high' (1);

(v) adessive forms of **place names** (168) – e.g. *Põltsamaal* 'at Põltsamaa';[32]

(vi) constructions expressing **state** (11) – e.g. *hirmul* 'afraid';

(vii) **verb government** (17) – certain verbs in Estonian require the verbal complement to be in the adessive case; the adessive complements of the following verbs were excluded from the analysis: *baseeruma/ põhinema/ rajanema millelgi* 'to be based on something' (8), *haarama kellelgi käest kinni* 'to take somebody's hand' (1), *kellelgi/ millelgi sabast kinni nabima* 'to catch somebody/something' (1), *mingil teemal juttu tulema/ kirjutama/ vestlema* 'to talk/write on a certain topic' (3), *laskma/ lubama kellelgi midagi teha* 'to let somebody do something' (3), *soovitama kellelgi midagi teha* 'to suggest somebody to do something' (1).

[32] Although it is possible to use the adposition *peal* with some of the place names in Estonian, these were left out of the present study because of meaning difference.

In addition to being an adposition, the word form *peal* also functions as an adverb in Estonian. Therefore, a great majority of the *peal*-tokens that were considered noninterchangeable involve cases when the word form *peal* is used as an adverb. The majority of the adpositional uses of *peal* that were excluded involve grammatical homonymy. The following specific types of ***peal*-constructions** were **excluded** from the analysis (the number in the brackets indicates the number of the specific types that is excluded):

(i) *peal*-form used as an **adverb** (355);

(ii) **fixed expressions** and **verb government** (288) – *ameti peal olema* 'to have a specific position' (8), *hea elu peal olema* 'to lead a good life' (2), *järje peal olema* 'to do ok in life' (1), *piiri peal olema* 'to be on the edge' (15), *pilku/silma peal hoidma* 'to keep an eye on' (61), *sina peal olema* 'to be informal with somebody' (5), *mingi asja peal väljas olema* 'to be eager about something' (30), *põhja peal* 'on the bottom of' (6), *vana rasva peal* 'from experience' (1), *viimase peal* 'extremely good' (8); fixed expressions with a noun expressing location (123), e.g. *linna peal* 'out and about the town' (lit. 'on town'); fixed expressions with vehicles (9), e.g. *bussi peal* 'on the bus'[33]; *millegi peal (ära) elama* 'to live on something' (4), *kellegi/millegi peal midagi katsetama* 'to try something on somebody/something' (16), *millegi peal üles kasvama* 'to grow up on something' (7), *kellegi peal välja elama* 'to take it out on somebody' (4), *millegi peal töötama* 'to work on something' (4);

(iii) *peal*-constructions that do not have an adessive alternative with the same meaning, i.e. cases of **grammatical homonymy** (436) – *koha peal* 'on the spot', cf. *kohal* 'above; present' (103), *maa peal* 'on earth/on ground', cf. *maal* 'in the country' (152), *pea peal*[34] 'on somebody's head', cf. *peal* 'on' (23), *peo peal* 'on somebody's palm', cf. *peol* 'at a party' (15), *poole peal* 'in the middle', cf. *poolel* 'on the side of' (30), *tee peal* 'on the road', cf. *teel* 'on somebody's way to' (73), *vee peal* 'on water', cf. *veel* 'still' (25).

Table 10 gives an overview of the types of adessive and *peal*-constructions that were excluded from the analysis, i.e. tokens coded 'noninterchangeable'. The

[33] The meaning of the adpositional phrase with a location or a vehicle noun followed by *peal* cannot be expressed with an adessive case; however, another locative case construction can be considered as a more appropriate alternative – the inessive case construction, cf. *linnas* [linn+INE] 'in town', *bussis* [bus+INE] 'in the bus'.

[34] Note that the adpositional phrase *pea peal* [pea+GEN peal], 'head' + 'on', can either mean that somebody is literally on his or her own head, there is something on somebody's head, or it can be used in a figurative sense of 'in a mess, upside down'. Compare also the word form *peal* which can either refer to the adposition or the adverb meaning 'on', or the adessive case form of the noun *pea* 'head', as in the sentence *Mu peal on valus* 'My head hurts'. This is a nice instance of how intricate grammatical homonymy and ambiguities can be that abound the Estonian language.

most frequent function of the adessive according to my analysis is to express temporal relations, followed closely by the function of expressing a logical subject as in the possessive construction. The other functions of the adessive are on bar with its function of expressing spatial relations (488 out of 3,680). Although as many as 1,309 *peal* tokens out of a total of 2,387 were coded as interchangeable with the adessive case, a relatively large number of *peal* tokens (436) are coded noninterchangeable due to reasons of grammatical homonymy. The number of *peal* tokens that are adverbs is 355 – this indicates, among other things, that *peal* used as an adposition is much more frequent in written present-day Estonian than *peal* used as an adverb.

Table 10. Type and number of adessive and *peal*-tokens coded as 'noninterchangeable' in multivariate corpus study

adessive tokens		***peal*-tokens**	
temporal	1,502	gram. homonymy	436
logical subject	1,451	adverb	355
adverb-like	373	fixed expressions	288
place names	168		
manner	158		
government	17		
state	11		
Total	**3,680**		**1,079**

Explanatory variables. Chapter 2 (specifically section 2.3) provides a detailed overview of the various explanatory variables selected for the present thesis, their operationalizations, the relevant examples, and the reasons for including them. I will not go into the description of these variables again, but will simply list here the specific variables included in the multivariate corpus study along with the specific levels (Table 11). The first column in Table 11 indicates to which group a specific variable belongs, the second column specifies the variable name, the third column indicates whether the variable is measured on a categorical (cat.) or interval (int.) scale, and the last column lists the variable levels. There are eleven semantic and nine morphosyntactic explanatory variables (20 in total) included in the multivariate corpus study. The majority of the variables are categorical, with the exception of length which is numeric, and have only two or three levels, with the exceptions of case form of Trajector (4 levels), and verb (5 levels).

Table 11. List of the explanatory variables in multivariate corpus study

Group	Variable name	Scale	Levels
semantic	type of relation btw LM & TR	cat.	abstract, spatial
	type of LM	cat.	place, thing
	type of TR	cat.	abstract, object
	animacy of LM	cat.	animate, inanimate
	animacy of TR	cat.	animate, inanimate
	number of LM	cat.	plural, singular
	number of TR	cat.	plural, singular
	mobility of LM	cat.	mobile, static
	mobility of TR	cat.	mobile, static
	relative sizes of TR & LM	cat.	convent., same, unconvent.
	verb	cat.	action, existence, motion, posture, no verb
morphosyntactic	length of LM phrase	int.	in syllables and words
	morphol. complexity of LM	cat.	simple, compound
	syntactic function of LM	cat.	adverbial, modifier
	word class of LM	cat.	noun, pronoun
	word class of TR	cat.	noun, pronoun, verb phrase
	case form of TR	cat.	nominative, partitive, other, not applicable
	clause type	cat.	main, subordinate
	position of LM phrase	cat.	initial, middle, final
	relative position of TR & LM	cat.	lm_tr, tr_lm

3.1.2 Corpus study 2: Distinctive Collexeme Analysis

Overview of the method. Distinctive collexeme analysis belongs to the set of corpus-linguistic methods referred to as **collostructional analysis** and developed by Gries and Stefanowitsch (Stefanowitsch and Gries 2003, 2005; Gries and Stefanowitsch 2004a, 2004b)[35]. The aim of collostructional analysis is to find out which lexical elements are typical for a given grammatical construction. This collocational information is taken as evidence for the semantics of the constructions. Distinctive collexeme analysis is used to study how two or more constructions differ from each other with respect to associated lexical material, where different collocational preferences are taken to reflect semantic differences between the constructions (Gries and Stefanowitsch 2004b: 97, Hilpert 2006: 243). Multiple distinctive collexeme analysis is an elaboration of distinctive collexeme analysis and it allows the comparison of

[35] The other two collostructional methods besides distinctive collexeme analysis are collexeme analysis (Stefanowitsch and Gries 2003) for determining which lexical items typically occur in a given slot in a single grammatical construction, and covarying-collexeme analysis (Gries and Stefanowitsch 2004a, Stefanowitsch and Gries 2005) for revealing dependencies between lexical items that occupy two different slots within the same construction.

more than two alternative constructions (Gries and Stefanowitsch 2004b). Future work on Estonian locative cases and adpositions may therefore consider other alternative adpositional constructions in addition to the adposition *peal* 'on' that can be used interchangeably with the adessive, e.g. the comparison of the adessive and the adpositions *peal* 'on', *otsas* 'on top of' and *küljes* 'attached to'.

Distinctive collexeme analysis, like other collostructional methods, is based on the idea of collocates – 'words that occur (with a frequency that is significantly above chance-level) in a given span around the node word' (Gries and Stefanowitsch 2004b: 100). Although such corpus-linguistic work based on collocates has led to interesting results, according to Gries and Stefanowitsch (2004b: 100) it has a limitation of disregarding syntactic structures. This in turn led to the implementation of collostructional[36] analysis methods that combine collocate analysis with syntactic and semantic structures in which words occur (Stefanowitsch and Gries 2003). Gries and Stefanowitsch (2004b: 101) acknowledge that the method of distinctive collexeme analysis is based on Church et al.'s (1991 cited in Gries and Stefanowitsch 2004b: 101) distinctive collocate analysis, but instead of looking at words the focus is on near-synonymous constructions.

Like many other corpus linguistic methods, distinctive collexeme analysis works on the basis of frequency counts and helps determine which elements occur more frequently in a construction than would be expected by chance. While general frequency counts also include the frequent items that are common to both constructions under study, distinctive collexeme analysis enables to abstract away from the common items and pinpoint those that are distinctive for each respective construction (Hilpert 2006: 243–250). The method identifies lexemes that occur significantly more often with one construction than with the other, and ranks these according to the degree to which they are distinctive. The calculation of the distinctiveness of a given collexeme is based on four frequencies: the lemma frequency of the collexeme in construction A, the lemma frequency of the collexeme in construction B, the frequencies of construction A and B (Gries and Stefanowitsch 2004b: 102). These frequency tables are then submitted to the Fisher exact test. The analysis is run by using the open source software R together with a script written by Gries (2007a). The output of a distinctive collexeme analysis is a pair of lists, which rank collocating items of the two constructions according to their collostructional strength.

Collostructional strength is a measure of the strength of the association and it is the log-transformed *p*-value from the one-tailed Fisher-Yates exact test. Gries and Stefanowitsch (2004b: 101) highlight some of the problems associated with using other measures than the Fisher exact test to determine association strengths and give reasons why the measure they propose is more

[36] Collostruction is a blend of construction and collocation (Gries and Stefanowitsch 2004b: 100).

suitable. The main argument for not using, for example, the *t*-score or the *z*-score, is that the use of these measures is not justified for studying natural language data which is hardly ever normally distributed and with a homogeneous variability (Gries and Stefanowitsch 2004b: 101). An exact test, on the other hand, does not make any distributional assumptions and does not require any particular sample size – it is also applicable in the many instances where given lexemes or constructions have a very low frequency (*ibid*.). Baayen (2011) provides some criticism on this measure of strength and discusses a number of alternative measures that can be used to calculate a lexeme's constructional preferences.[37] Although he concludes that all the measures are "correlated and useful as measures of collexeme strength" (Baayen 2011: 320), they should all be fully grounded in learning theory as well. It is not sufficient for corpus linguists to just explore the distribution of elements – the descriptive approach has a lot to profit from computational modelling. It is also for this reason why a purely descriptive distributional account (i.e. the distinctive collexeme analysis and the univariate results of the multivariate corpus study) are supplemented by regression modelling in the present study.

Limitations of the method. One of the major limitations of distinctive collexeme analysis is that it does not take into account the overall corpus frequencies of the lexemes that occur in the two constructions – it merely highlights differences and does not characterise the two constructions per se (Hilpert 2006: 250). As suggested by Hilpert (2006: 250) this limitation should be alleviated by the application of two separate collexeme analyses. In this case, one separate collexeme analysis should be run with the adessive and the other with *peal*. Separate collexeme analyses take into account the relative frequencies of the lexical items themselves as well. However, this limitation can be disregarded on the grounds that the main interest of the present thesis is in finding out which lexemes distinguish between the two specific constructions rather than how the two constructions contrast to other (locative) constructions in general. Using separate collexeme analyses, in turn, brings along another disadvantage, associated specifically to this method – it relies heavily on tagged corpora to enable the exhaustive retrieval of the constructions and the lexical items. The availability of such data is not only problematic for diachronic studies (Hilpert 2006: 250), but also for heavily inflected languages like Estonian. As was already pointed out in the previous section, the size of the Morphologically Disambiguated Corpus of Estonian does not allow, at least in its current stage, the large-scale quantitative analysis of the adessive and *peal*-constructions as desired. A collexeme analysis based on an untagged corpus involves much work as pertains to retrieving the items and separating the wheat from the chaff, i.e. including only the examples that are truly relevant for the study. Distinctive collexeme analysis, in comparison, has the advantage that it

[37] See also the discussion on frequency and the cognitive underpinnings in Chapter 1 above and Schmid (2010) for criticism on this and other measures of frequency.

can be used with untagged corpora. The effort is still there, but in magnitudes smaller than that required for separate collexeme analyses. As Hilpert (2006: 251) points out "all necessary data points are contained in the concordances of the investigated construction".

In assessing the usefulness of this method, Gries and Stefanowitsch (2004b: 118–123) address two major issues – can the method be used to making above-chance level predictions about which member of an alternating pair will be chosen for a particular lexeme and is the method overly sensitive, i.e. does it produce significantly distinctive collexemes even for alternating pairs where this would not be plausibly expected. They give a positive answer to the first question and a negative one to the second, showing that the distinctive collexeme analysis does indeed allow the prediction of the choice of construction above chance, even if it cannot predict constructional choices in all cases (Gries and Stefanowitsch 2004b: 120–121), and that if you apply the method to an alternating pair where one would not expect any distinctive collexemes, one would find that there are almost no distinctive collexemes as predicted (Gries and Stefanowitsch 2004b: 122). Gries and Stefanowitsch (2004b: 122) therefore suggest not to throw out the baby with bathwater "by prematurely dispensing with a technique whose other merits are quite obvious". It was decided, therefore, to apply this method for the present research question as well and to see if and how well it works for an alternating pair of the kind represented by the Estonian adessive and adposition *peal* 'on'; at the same time bearing in mind all the aforementioned limitations of this method.

The data sample. The data for distinctive collexeme analysis with Estonian adessive and *peal*-constructions comes from the fiction sub-corpus of the Balanced Corpus of Estonian (2008; size 5 million words). Collostructional analysis relies heavily on tagged corpora, but as we already saw above in the previous section, the size of the Morphologically Disambiguated Corpus of Estonian is not sufficient for a large-scale analysis of the two constructions. Hence it was decided to use the 5-million fiction sub-corpus of BCE. Following Gries and Stefanowitsch (2004b: 105) it was decided to include only those lexemes that occur at least once in each construction. This approach enabled to overcome the obstacle of not having a large enough tagged corpus available for the analysis. It was possible to extract a list of all lexemes[38] occurring with the adposition *peal* 'on' in the fiction sub-corpus of BCE using a Perl script. Each lexeme in this list was then individually entered into the corpus query in the form of the adessive, both in singular and plural – if the noun did not occur in the adessive construction, it was excluded from further analysis. As we have seen in countless previous occasions already, the adessive frequently fulfils other functions besides the locative function where the two constructions are interchangeable. It was therefore necessary to go through all of the adessive case constructions manually and to confirm whether an instance was indeed

[38] Only full nouns were included in the analysis.

interchangeable with the *peal*-construction. The decision of interchangeability was based on the same criteria as described above for the multivariate corpus study, excluding thus all of the adessive constructions where it expressed the logical subject or cases of verbal government. The most important criterion was that the overall meaning of the adessive construction would not be changed if it were used with the adposition *peal* 'on' instead. The total number of the adessive case constructions retrieved before manual clean-up was 9,087. After manual post-editing, slightly over 1,000 instances were discarded and the final analysis included 7,979 instances of the adessive construction in total. The total number of the *peal*-constructions was 703. These general results mirror the results of the multivariate corpus study in that the adessive construction is about 10 times more frequent than the adpositional *peal*-construction expressing location in present-day written Estonian.

The final list of lexemes included 182 different nouns. The frequencies for each of the 182 words were cross-tabulated with the frequencies of the two constructions as shown in Table 12. From Table 12 it can be seen that the noun *laud* 'table' occurs 345 times with the adessive and 47 times with the adposition *peal*. Presented with such raw frequency lists we are left to assume that the noun *laud* 'table' significantly more often occurs in the case construction than in the adpositional construction. However, including the frequency information of other nouns occurring in these two constructions allows us to evaluate these results in a more reliable way. More importantly, this technique factors in the fact that the adessive construction is much more frequent than the adpositional construction in the first place. If we do not take this information into account, we might be left with the understanding that no lexeme is distinctive for the adpositional construction – a result which is clearly not valid, as the results presented in Chapter 4 demonstrate.

Table 12. Distribution of four frequent collexemes with adessive and adposition *peal* 'on' in the fiction sub-corpus of BCE

Collexeme	Frequency with adessive	Frequency with *peal* 'on'	Row totals
laud 'table'	345	47	392
sein 'wall'	269	17	286
pilt 'picture'	236	35	271
uks 'door'	226	24	250
...	...	...	...
Column totals	7,979	703	8,682

Predictions. The reason why distinctive collexeme analysis is potentially useful for the study of the alternation between Estonian locative cases and adpositions lies in the fact that there are a number of nouns that occur in both constructions. This would lead us to assume, as has been in fact assumed in Estonian reference

grammars, that the two constructions are semantically equivalent. However, there may be subtle differences between the two constructions in terms of the semantic restrictions they exhibit. Establishing the nouns that exhibit a strong preference for one locative construction (e.g. the adessive) as opposed to the other (e.g. the adposition *peal* 'on') may show that locative cases prefer one type of Landmarks and the adpositions another type. This issue is addressed also in the multivariate corpus study and it is interesting to see to what extent there is converging and/or diverging evidence.

Similarly to the discussion on the English dative alternation in Gries and Stefanowitsch (2004b: 104–105), it can be predicted according to the approaches that disregard semantics as a valid variable that the two constructions (adessive and *peal*-construction in the present case) should not differ at all with respect to preferred lexemes. The semantic approach, e.g. a Cognitive Grammar (Langacker 1987, 2008) or Construction Grammar (Goldberg 1995, 2006) approach, predicts that such differences do exist. Based on previous research on cases and adpositions, and based on the results of multivariate corpus study, it can be predicted that the adessive construction prefers lexemes that denote static places and the adpositional construction lexemes that denote small, movable things.

3.2 Experimental studies

In Chapter 1 it was stressed that corpus linguistic work should be supplemented with experimental work and that methodological pluralism is the preferred choice in empirical linguistic research. This section describes the experimental designs opted for in the dissertation and lists some of the reasons why these particular designs were chosen. Experimental research is seen as the other side of statistical inference – that of hypothesis testing – and it is therefore seen as complementing the two corpus studies described in the previous section. As corpus linguistic work preceded the linguistic experiments, the two tasks test some of the major findings of the two corpus studies – the semantic variable 'type of Landmark' and the morphosyntactic variable 'length of the Landmark phrase'. These two variables were selected because they represent different groups of variables studied in the corpus analysis and because they had the strongest effect size within their respective categories. In addition, the morphosyntactic variable word order (relative position between Trajector and Landmark) was examined based on my own intuition as a native speaker of Estonian. The two linguistic experiments included in this study are both **rating tasks**[39]. The first rating task is referred to as the **picture rating task** and the

[39] I am using the term "rating task" rather than "acceptability judgement task" due to the specific instructions that were given to the subjects, i.e. they were not asked to judge the acceptability of the sentences, but they were asked to judge how likely it is they would use the sentences themselves (see sections 3.2.1 and 3.2.2 for details about the exact instructions). However, since the general design of both tasks is similar to acceptability

second as **sentence rating task**. The reasons why rating tasks rather than other types of experimental designs were used have to do with the main aim of the thesis, i.e. to discuss data from multiple sources.

Previous experimental work on this topic includes a pioneering forced-choice task and a semi-open production task reported in Klavan et al. (2011). As mentioned in section 1.3.5 above, the results of such experimental designs are shown to converge fully with the results of corpus studies (Arppe and Järvikivi 2007). The nature of the data obtained with a production task is in essence the same as that of the data obtained from a corpus. Clearly, the data may differ as to whether the production is spoken or written, but for a researcher, the tokens produced by language users are like extractions from an already compiled corpus. Therefore, a similar methodological approach can be taken – estimation of the population parameters on the basis of a given sample rather than specific hypothesis testing. Arppe and Järvikivi (2007: 150–151) also point out that a forced-choice task reflects actual usage situations and is about linguistic performance in production, making the data obtained with this type of linguistic experiment again similar to that of a corpus analysis[40]. A rating task, on the other hand, reflects what is considered possible or appropriate and it is about introspection.

Dąbrowska (2008: 395) makes a similar point by claiming that many linguists would argue that a judgement task "provides more useful evidence about the nature of speakers' underlying linguistic representations, or 'competence' [...] and, hence, perhaps will be less likely to be dismissed as 'mere performance'". However, she (Dąbrowska 2008: 395) goes on to acknowledge that judging a sentence is a type of performance and is likely to be influenced by a variety of factors such as complexity, fatigue, mode of presentation, etc.[41] Similarly to Dąbrowska (2008), I also believe that the solution is not to give up studying speakers' judgements experimentally, but to try and control as many

---

judgement task, the practical considerations of the latter hold for the two rating tasks as well. Arppe and Järvikivi (2007: 155, Ftn. 2) point out that experimental judgements may concern acceptability, naturalness, grammaticality, ungrammaticality, well-formedness, correctness, interpretability, ill-formedness, probability of occurrence or preference of choice. From this list, the ratings in the studies reported here conform best to probability of occurrence. Similarly to Arppe and Järvikivi (2007: 142), the ratings pertain to acceptability rather than grammaticality since the analysis of the corpus showed that both constructions can be considered well-formed (i.e. grammatical) in the contexts provided in the two tasks. See also Dąbrowska (2010: 3–5) for a more recent discussion on the distinction between grammaticality and acceptability.

[40] Cf. Schütze and Sprouse (to appear) who consider a forced-choice task as one type of judgement task along with magnitude estimations and Likert scaling and for whom the cognitive task the subjects perform is identical across all of these measures. The data yielded by each task is therefore very likely to be similar.

[41] See also the discussion in Schütze (1996: 98–169), one of the first researchers to address these issues, for various subject-related (e.g. handedness, literacy, and education) as well as task-related factors (e.g. instructions, order of presentation, modality, frequency and complexity of the stimulus) that influence acceptability judgements.

confounding factors as possible. In addition, the employment of adequate statistical techniques, such as ordinal mixed-effects logistic regression or repeated measures analysis of variance (see section 3.4 below for overviews), enables to come to grips with the relatively noisy and volatile nature of individual judgements. However, it is still necessary to keep in mind all the practical aspects of the specific experimental designs when interpreting the results as these are likely to influence the results. A brief discussion on some of the most relevant practical aspects of the two rating tasks reported here is given below.

Another reason for choosing the experimental design of a rating task over other designs is that it permits the discussion of **how frequency relates to acceptability**. As was already seen in Chapter 1, recent linguistic studies that employ both corpus as well as experimental methods frequently ask the question whether frequency is mirrored by acceptability (Arppe and Järvikivi 2007; Divjak 2008, in progress; Schmid 2010). A plausible prediction is that language users give the more frequent linguistic structure in a particular context a significantly higher rating than to the less frequent one. However, there exists a vast amount of research indicating the opposite – the rareness or infrequency of a particular structure in a corpus is not necessarily associated with a lower rating (e.g. Arppe and Järvikivi 2007; Divjak 2008, in progress; Schmid 2010). As will become evident in the next two chapters, the two rating tasks present diverging results. While the picture rating task reflects, in general, the corpus results where higher frequency maps onto higher rating, the same conclusion cannot be drawn for the sentence rating task. Such diverging results confirm the conclusion made by other researchers that we should be careful about reading too much into frequency information as provided by corpus linguistic research and that further studies are needed to determine how exactly frequency relates to acceptability. I will take up this discussion again in Chapter 6.

A further advantage of rating tasks over other experimental designs is that they allow access to more fine-grained measures than simple frequency counts of the tokens in a production task or a simple binary "yes" or "no" response in a forced-choice task. The two latter types of tasks are designed to detect qualitative differences between conditions, but they do not say anything about the size of the difference, something that rating tasks are able to provide (Schütze and Sprouse to appear: 6). The bone of contention, however, is whether to treat the rating participants give as a score on an **interval scale** or a score on an **ordinal scale**, i.e. as a quantitative or a qualitative measurement.[42] There are arguments in the literature in favour and against both approaches. It is frequently pointed out that for acceptability judgement experiments using Likert

[42] The questions concerning the measurement scales become crucially important when choosing the appropriate statistical technique to run the analysis – the choice very much depends on the nature of the data and each technique has its own assumptions (see sections 3.3 and 3.4 below for the assumptions of the techniques employed in this study).

scaling[43], the scales are ordinal rather than interval since there is no guarantee that the intervals are equally spaced (cf. e.g. Carlson and Hill 2007: 256, Schütze and Sprouse to appear: 8). For example, on a scale from 1 to 7, the difference between values 2 and 3 may or may not be the same as the difference between values 6 and 7. Another problem with using scales is that subjects tend to use only the endpoints or middle values of the scale compressing the scale in effect into only three response categories; people may, furthermore, use the same scale differently (Carlson and Hill 2007: 257). However, there are statistical approaches that help to mitigate the effect of such scale bias, such as mixed effects modelling (see section 3.4 below).

Since we do not know whether or not the assumptions of using a multivariate technique designed for interval data (e.g. such parametric techniques as repeated measures ANOVA, see below for details) are violated or not, it would be safer to employ techniques designed specifically for ordinal data (e.g. ordinal logistic regression, see below for details). Nevertheless, some researchers have employed ANOVA to analyse the results of such (psycho)linguistic experiments where people provide ratings on a Likert-type scale (e.g. Carlson-Radvansky and Tang 2000; Carlson and Logan 2001; Coventry et al. 2001; Dąbrowska 2008, 2010). The basic argument is that since the results are reliable, the use of multivariate techniques is justified. As Rencher (2002: 2) points out: "[s]ome authors warn against applying the common multivariate techniques to data for which the measurement scale is not interval or ratio[, i]t has been found, however, that many multivariate techniques give reliable results when applied to ordinal data." Furthermore, Schütze and Sprouse (to appear: 19) simply disregard the arguments against considering the responses on a Likert scale as interval (i.e. numeric) and for employing parametric statistical tests on the grounds that the field "has decided (consciously or not) that it is willing to tolerate the potential consequences of the violations of parametric tests" since these tests involve assumptions that are rarely met in psychological research anyway.

As there are arguments in favour and against using the ratings as either interval or ordinal, and since the aim of the thesis is to compare different methodologies, it is thought best to take a look at the results from both perspectives in the present study. Reference is also made to other experimental work besides rating tasks on the same topic. Specifically the forced choice task and the semi-open production task reported in Klavan et al. (2011) are referred to. A common disadvantage of rating tasks (and experimental tasks in general) is that they permit the study of only a selection of variables. Langacker (1987:

[43] Magnitude estimation (ME) is frequently seen as a valid alternative experimental design to Likert scaling (LS) (Cowart 1997, Keller 2000, Schütze and Sprouse to appear). However, it was decided not to use it for the present study because I have no experience with ME and it has been shown that the data obtained with such a techniques is ordinal as with scale tasks: "participants may be treating the ME task as a type of LS task" (Schütze and Sprouse to appear: 9). Weskott and Fanselow (2009, 2011) draw a similar conclusion and claim that ME data are not more informative than categorical data or scalar judgements.

37), like many others, rightly doubts the context-independent nature of judgements and claims that judgements about sentences are always made relative to real or imagined contexts. He also argues that judgements are influenced by various factors (Langacker 1987: 36). Sound as such criticism is, and it naturally pertains not only to rating tasks but to other types of experimental designs as well, it is precisely the aim of linguistic experiments to control for as many variables as possible in order to test the influence of one or more specific variables. Another disadvantage of rating tasks is that because ratings are in a way an artificial measure that does not necessarily map onto actual use, corpus data is included in the discussion. Again the reason for including these different studies in the discussion is to see how much there is converging or diverging evidence. At the same time, we should keep in mind the concerns voiced in Chapter 1 about comparing data from different sources.

Some practical considerations pertaining to the rating tasks. Since acceptability judgement tasks are one of the most widely used data sources in linguistic research[44] (Schütze 1996, Cowart 1997, Keller 2000, Featherston 2005b, Myers 2009, Dąbrowska 2010, Sprouse 2011, Schütze and Sprouse to appear), the few textbooks and practical papers that do exist on doing linguistic experiments deal precisely with the design issues of such tasks. Excellent practical sources are, among others, Schütze (1996), Cowart (1997), Myers (2009) and Schütze and Sprouse (to appear). Following is a discussion of some of the specific task-related factors that need to be born in mind when interpreting the results presented in Chapter 5. One of the first things that needs to be decided is **choosing the response scale**, which involves determining the number of response categories, whether there should be a midpoint value, and determining the labels to use and the number of response categories to be labelled. The two rating tasks use a 5-point and 7-point scale with the two endpoints labelled. The participants were free to use a midpoint value, but this is not labelled. In addition, **anchoring** is used to help ensure that participants use the scale the same way. That means that for both rating tasks, anchor items are provided for the lowest and highest point on the scale.

It was decided to use a 5-point scale for one experiment and a 7-point scale for the other due to methodological reasons. Since one subset of my research interests includes the methodological underpinnings of running linguistic experiments, future work in this direction may benefit from comparing the nature of the results provide for the two experiments vis-à-vis the response scales applied in the experiments. The results presented in Chapter 5 indicate that in both experiments the categories 5 and 7 were the most frequent ratings with the other rating categories being more or less equally distributed.

[44] The reasons why judgement experiments enjoy such priviledged status vary; most common are the following: they have a long history in syntactic research, they are relatively easy to design and conduct (vis-à-vis other experimental designs), no special apparatus is required, etc.

Care should be taken in designing the **instructions**. Several issues are at stake here: first, to make it clear to the participants that the whole range of the scale should be used; second, to make it clear to the participants what it is that they are supposed to do. Schütze (1996: 132) points out that "even subjects who are supposedly experts on language cannot be expected to know what linguists mean by *grammaticality* (or *acceptable*, for that matter)". This may lead to the situation where every participant takes their own interpretation and the results are meaningless. The specific instructions of the two rating tasks are given in sections 3.2.1 and 3.2.2. It should be pointed out that a different wording from a majority of acceptability judgement tasks is used in the two rating tasks. Rather than asking the participants to rate the grammaticality or acceptability of the sentences, it was decided to ask them to rate how probable it is that they themselves would use these sentences. The reason has to do with the above-mentioned difficulty in determining what exactly is meant by the notions of grammaticality and acceptability and to make the rating measure less artificial and conform more to actual language use. It would be, however, interesting to see what results will be given should the wording in the instructions be changed. This line of research is, albeit, left open for the future.

A further important point is the **modality** and register of presentation. Schütze (1996: 147) reports that there is no modality effect which is at odds with the belief that judgement criteria are much stricter for written language than for spoken language. Both of the rating tasks reported here used the written mode of presentation (cf. the corpus data which also comes from the written mode). In the picture rating task, the sentences to be rated are accompanied by pictures; the second rating task involves only the sentences. It is important to bear in mind that both of the tasks were conducted via the Internet[45] and not on paper. As with any decision related to a design issue, there are both advantages and disadvantages. The advantage of using a task designed on the Internet is that the participants are not able to go back and change their responses, since once the sentence was rated, a new page appeared with the next sentence and there was no going-back button. Distributing the questionnaires via the Internet also enabled to recruit a larger number of participants (73 and 103 respectively)[46]. The disadvantage is that the participants were free to complete the task whenever it suited them best and there is no guarantee that they did not consult with anyone else during the task. Again, it would be interesting to see if and how the results change were the tasks presented on paper, something left for future research. Although the eFormular tool used for designing the tasks allows access to response times and it is possible to see how long did it take for each participant to rate each sentence, any such rating task can of course only

[45] The two rating tasks were designed and distributed using the online questionnaire tool eFormular, available at http://www.eformular.com.

[46] The call for participation was sent out via different social networks. The subjects who participated in the picture rating task were not the same as those who participated in the sentence rating task.

provide indirect evidence about sentence processing. As such, the rating tasks reported here are off-line rather than on-line experiments.

Another set of practical considerations relates to the selection and number of **stimuli**. A number of rules of thumb are applicable: sentence order should be controlled for, e.g. by **randomization**; the stimuli should be controlled for lexical content, meaning, parsability, frequency, length, and structure (Schütze 1996: 150–168); it is desirable to create multiple lexicalization of each condition (Schütze and Sprouse to appear: 13); etc. The two rating tasks reported here both used randomized ordering and multiple versions with a different order of the same task. The stimuli in both tasks were controlled for the abovementioned variables and four lexicalizations were created for each condition. Although in an ideal experimental situation each participant rates only one sentence per trial and does not see the same lexicalization of related conditions, in the two rating tasks all of the participants saw all of the lexicalizations of the conditions. In the sentence rating task, all participants rated all of the 40 sentences; in the picture rating task, the stimulus set was divided in half in order to make the questionnaire shorter and to avoid fatigue effects. According to Schütze (1996: 135–136), **repetition** may have three possible outcomes: first, ratings might become more lenient due to the general psychological phenomenon of habituation; second, ratings might become stringent as participants have repeated exposure to the syntactic or semantic properties of the sentences; and third, repetition might just increase participants' confidence in their original judgements – there is polarization, where good sentences get better and bad sentences worse. The use of multivariate statistical techniques like repeated measures ANOVA and mixed-effects regression in analysing the results alleviates these design issues.

It is common and beneficial to use **fillers**, i.e. sentences that are not related to the research question, in experiments (Schütze and Sprouse to appear: 14). Fillers mostly serve the purpose of confusing the participants, i.e. reducing the chances of participants guessing the aim of the experiment and becoming aware of the particular sentence type that is being tested, which in turn may trigger conscious response strategies. Despite this advice, it was decided not to use fillers for the two rating tasks reported here. First and foremost, it was not the aim to "hide" the main aim of the experiment from the participants and it was, indeed, the goal to encourage the participants to work out a conscious response strategy in differentiating between the two locative constructions. However, the comments received after the experiment clearly demonstrated that the participants did not guess the different conditions, i.e. they were only aware that the two different locative constructions were tested, but not the type of Landmark or its combination with word order or length. The second reason for excluding fillers had to do with the length of the tasks. As all participants saw all of the sentences, the questionnaires were already relatively long.

The inevitable conclusion is that there are costs and benefits to every experimental design and every design issue. As long as one is aware of these issues and finds at least some ways to alleviate the potential biases in the data,

adequate conclusions can be drawn about the linguistic phenomenon under study. The next two sections will give a detailed description of the two rating tasks. The main aim in designing these tasks is to test some of the hypotheses arrived at after examining the two corpora. The results show that a variety of morphosyntactic and semantic variables play a role in the alternation between the two Estonian locative constructions. Among the significant variables are length of the Landmark phrase (the longer the phrase, the more frequent the use of the synthetic adessive case) and the type of Landmark (small, manipulable things prefer the analytic adpositional construction; larger, static places the synthetic case construction). In order to address the question of what role these variables play individually, the two rating tasks were designed. The picture rating task looks at the type of Landmark and the sentence rating task at the length of the Landmark phrase. In addition, both tasks investigate how the position of the Landmark phrase relative to the Trajector phrase contributes to participants' judgements. Since the first rating task addresses a semantic question, pictures are used to accompany the sentences.

3.2.1 Rating task I: Picture rating task

In the first experimental study, native speakers of Estonian completed a picture rating task in which they were asked to rate the probability with which they themselves would use the sentences with both the adessive case and the adposition *peal* 'on' to describe the pictures presented in the task.

Predictions. The first rating task tests the influence of two variables: type of Landmark and the relative word order sequence between the Landmark and Trajector phrase. A more detailed discussion of the variables is given above in section 2.3. Based on this discussion and the results of the corpus studies, the following predictions are made:

> Prediction 1: Adessive sentences with small manipulable things as Landmarks should receive lower ratings than the corresponding sentences with the adposition *peal* 'on'.
>
> Prediction 2: Adessive sentences with body parts as Landmarks should receive higher ratings than the corresponding sentences with the adposition *peal* 'on'.
>
> Prediction 3: Adessive sentences with furniture as Landmarks should receive lower ratings than the corresponding sentences with the adposition *peal* 'on'.
>
> Prediction 4: Adessive sentences with building parts as Landmarks should receive higher ratings than the corresponding sentences with the adposition *peal* 'on'.

Prediction 5: Adessive sentences with places as Landmarks should receive higher ratings than the corresponding sentences with the adposition *peal* 'on'.

Prediction 6: Adessive sentences where the Landmark phrase precedes the Trajector phrase should receive higher ratings than the corresponding sentences with the adposition *peal* 'on'.

Prediction 7: Adessive sentences where the Landmark phrase follows the Trajector phrase should receive lower ratings than the corresponding sentences with the adposition *peal* 'on'.

Experimental conditions. The sentences were with two different word order sequences (WO), five different sets of Landmark types (LM), and two types of constructions (CX) eliciting thus the following 20 experimental conditions given in Table 13. The table gives examples of each type of sentence and the English translations; a complete list of all experimental sentences used in the study is given in Appendix 1. For example, the first experimental condition (the row titled '1.' In Table 13) represents adessive sentences with a body part as a Landmark and the Landmark phrase preceding the Trajector phrase (Lm1, i.e. the Landmark comes first); the second experimental condition represents a corresponding *peal*-sentences.

Table 13. Example stimuli used in picture rating task

	Condition	Example	LM	WO	CX
1.	bodypart_lm1_ade	Õlal on kast.	bodypart	Lm1	ade
2.	bodypart_lm1_peal	Õla peal on kast.	bodypart	Lm1	peal
		'On the shoulder is a box'			
3.	bodypart_lm2_ade	Kast on õlal.	bodypart	Lm2	ade
4.	bodypart_lm2_peal	Kast on õla peal.	bodypart	Lm2	peal
		'The box is on the shoulder'			
5.	furniture_lm1_ade	Laual on tass.	furniture	Lm1	ade
6.	furniture_lm1_peal	Laua peal on tass.	furniture	Lm1	peal
		'On the table is a cup'			
7.	furniture_lm2_ade	Tass on laual.	furniture	Lm2	ade
8.	furniture_lm2_peal	Tass on laua peal.	furniture	Lm2	peal
		'The cup is on the table'			
9.	building_lm1_ade	Põrandal on vaip.	building	Lm1	ade
10.	building_lm1_peal	Põranda peal on vaip.	building	Lm1	peal
		'On the floor is a carpet'			
11.	building_lm2_ade	Vaip on põrandal.	building	Lm2	ade
12.	building_lm2_peal	Vaip on põranda peal.	building	Lm2	peal
		'The carpet is on the floor'			
13.	place_lm1_ade	Põllul on traktor.	place	Lm1	ade
14.	place_lm1_peal	Põllu peal on traktor.	place	Lm1	peal
		'On the field is a tractor'			

	Condition	Example	LM	WO	CX
15.	place_lm2_ade	Traktor on põllul.	place	Lm2	ade
16.	place_lm2_peal	Traktor on põllu peal.	place	Lm2	peal
		'The tractor is on the field'			
17.	thing_lm1_ade	Raamatul on kell.	thing	Lm1	ade
18.	thing_lm1_peal	Raamatu peal on kell.	thing	Lm1	peal
		'On the book is a watch'			
19.	thing_lm2_ade	Kell on raamatul.	thing	Lm2	ade
20.	thing_lm2_peal	Kell on raamatu peal.	thing	Lm2	peal
		'The watch is on the book'			

Stimuli: Experimental sentences. The experimental sentences were constructed by combining a "sentence stub" with a completion consisting of a Landmark phrase, a Trajector phrase and a verb. The Landmark phrase either followed the Trajector phrase, as in example 55, or preceded it, as in example 56.

(55) Trajector phrase Verb Landmark phrase

(56) Landmark phrase Verb Trajector phrase

The completion for the Landmark phrase consisted of a noun inflected for the adessive case (example 57) or a noun used in the postpositional phrase with the adposition *peal* 'on' (example 58):

(57) *laual*
table.SG.AD
'on the table'

(58) *laua* *peal*
table.SG.GEN on.AD
'on the table'

The completion of the Trajector phrase also consisted of a one word noun phrase (e.g. *tass* 'cup') and all experimental sentences contained the verb *olema* 'be' in the present tense, 3rd person singular form. Thus, all the experimental sentences were 3 (sentences with the adessive case) or 4 words (sentences with the adposition *peal* 'on') long and contained three elements.

Constructing the questionnaire. In total, 20 different Landmark and Trajector pairs were used to construct the experimental sentences with 4 Landmark-Trajector pairs per each of the five Landmark types. The completion for conditions 1-4 (body parts as Landmarks) consisted of these Landmark-Trajector pairs: *shoulder-box, back-rock, cheek-flag, nose-Band-Aid*; the completion for conditions 5-8 (furniture as Landmarks) consisted of the following pairs: *bed-rose, table-cup, chair-bag, bench-suitcase*; the completion for conditions 9-12 (parts of a building as Landmark) included the following

Landmark-Trajector pairs: *roof-aerial, wall-painting, floor-carpet, stairs-basket*; the completion for conditions 13-16 (places as Landmarks) consisted of these Landmark-Trajector pairs: *meadow-wagon, street-bus, shore-tree, field-tractor*; and last, the completion for conditions 17-20 (small manipulable things as Landmarks) included the following Landmark-Trajector pairs: *book-clock, rock-leaf, clothesline-towel, ladder-tin of paint*.

For each of the 20 Landmark and Trajector pairs four sentences were constructed – two sentences were such that the Landmark phrase preceded the Trajector phrase (one with the adessive case construction and one with the adposition *peal* 'on') and two sentences were such that the Landmark phrase followed the Trajector phrase (again, one sentence was with the adessive case and the other with the adposition *peal* 'on'). Thus, in total, there were 20 sentences with each of the five types of Landmark, making it a 5 (type of Landmark: bodypart, furniture, building, place, thing) x 2 (word order: landmark first, landmark second) x 2 (locative construction: adessive, *peal*) cross-factorial design. In addition, because there were four different lexicalizations of the Landmark selected for each of the 5 types of Landmark, there were 80 experimental sentences in total.

It was decided to make two versions of the study, since otherwise it would have been too tedious and repetitive for one subject to rate all of the 80 sentences. Both of the versions consisted of 40 experimental sentences. The experimental sentences were divided between the two versions so that all of the 20 Landmark-Trajector pairs were included in both of the versions. In each of the versions, there were two sentences with each Landmark-Trajector pair with word order varying between the two versions. Table 14 illustrates the way the experimental sentences were assigned to one of the two versions. In addition, there were two randomizations of the two versions of stimuli. A full list of the sentences used in both versions of the experiment is given in Appendix 1.

Table 14. Division of experimental sentences between the two versions of picture rating task

Version 1	Version 2
1) Tass on laual.	3) Laual on tass.
2) Tass on laua peal.	4) Laua peal on tass.
"The cup is on the table."	"On the table is a cup."
5) Põrandal on vaip.	7) Vaip on põrandal.
6) Põranda peal on vaip.	8) Vaip on põranda peal.
"On the floor is a carpet."	"The carpet is on the floor."

The task also included 20 pictures that were selected using the Google image search to represent the 20 Landmark-Trajector pairs corresponding to the experimental sentences. The pictures were real-life photos and they were formatted so that they were more or less of the same size and quality. A full list of the images is given in Appendix 1 together with sentence stimuli. In each of the two versions of the study, each picture appeared twice – once with the adessive case sentence and once with the sentence with the adposition *peal* 'on'. The two pictures below accompanied the eight sentences given in Table 14 above; the first picture accompanies the sentences 1–4 and the second picture the sentences 5–8:

Picture 1: cup on the table

Picture 2: carpet on the floor

Participants. 68 native speakers of Estonian aged between 17 and 46 (mean age 26 years) participated in the task. 42 of the subjects were women and 26 men. 45 of the subjects came from South Estonia, 14 from North Estonia, and 9 from other parts of Estonia. 34 of the subjects marked "university" as their highest school that they have finished, 23 "secondary school", 3 "applied higher education institution", 4 "vocational school" and 4 "basic school".

Procedure. Participants were asked to complete a picture rating task via the Internet using the eFormular tool and were given the following instructions:

> The rating task is part of a study on Estonian locative cases and spatial adpositions. It takes about 10 minutes to complete the task. The task is anonymous. It is not an intelligence test or a grammar test; there are no right or wrong answers. I am only interested in the language use.
> Your task is to rate on a scale from 1 (not very probable) to 5 (very probable) how probable it is that you yourself would use the sentences given in the task to describe the pictures presented together with the sentence. Read the sentences carefully, but do not spend too much time on it – I am interested in your initial reaction.

The instructions were followed by two example sentences and corresponding pictures for which ratings were provided. The first example sentence was given a rating of 1 and the second a rating of 5. These examples were provided in order to anchor the participants' ratings. The picture rating task was distributed electronically. Participants were randomly assigned to one of the two versions of the study, with 31 completing the first version and 37 the second version of the picture rating task.

3.2.2 Rating task 2: Sentence rating task

In the second experimental study, a different sample of native speakers of Estonian completed a written rating task in which they were asked to rate the probability with which they themselves would use the sentences given in the task. Since the sentence rating task tested the morphosyntactic variables, no pictures were included in this task. The participants only saw the sentences.

Predictions. The sentence rating task tests the influence of two variables: length of the Landmark phrase and the relative word order sequence between the Landmark and Trajector phrase. A more detailed discussion of the variables is given above in section 2.3. Based on this discussion and the results of the corpus studies, the following predictions are made:

> Prediction 1: Adessive sentences with long Landmark phrases should receive higher ratings than the corresponding sentences with the adposition *peal* 'on'.
>
> Prediction 2: Adessive sentences with short Landmark phrases should receive lower ratings than the corresponding sentences with the adposition *peal* 'on'.
>
> Prediction 3: Adessive sentences where the Landmark phrase precedes the Trajector phrase should receive higher ratings than the corresponding sentences with the adposition *peal* 'on'.
>
> Prediction 4: Adessive sentences where the Landmark phrase follows the Trajector phrase should receive lower ratings than the corresponding sentences with the adposition *peal* 'on'.

Experimental conditions. Similarly to the picture rating task described in the previous section, native speakers of Estonian completed a written questionnaire in which they were asked to rate the sentences with both the adessive case and the adposition *peal* 'on' with different word order sequence and with Landmark phrases of different lengths. There were 8 experimental conditions:

1. Adessive sentences with short Landmark phrases with the Landmark phrase preceding the Trajector phrase (ade_short_lm1)
2. *Peal*-sentences with short Landmark phrases with the Landmark phrase preceding the Trajector phrase (peal_short_lm1)
3. Adessive sentences with short Landmark phrases with the Landmark phrase following the Trajector phrase (ade_short_lm2)
4. *Peal*-sentences with short Landmark phrases with the Landmark phrase following the Trajector phrase (peal_short_lm2)
5. Adessive sentences with long Landmark phrases with the Landmark phrase preceding the Trajector phrase (ade_long_lm1)
6. *Peal*-sentences with long Landmark phrases with the Landmark phrase preceding the Trajector phrase (peal_long_lm1)
7. Adessive sentences with long Landmark phrases with the Landmark phrase following the Trajector phrase (ade_long_lm2)
8. *Peal*-sentences with long Landmark phrases with the Landmark phrase following the Trajector phrase (peal_long_lm2)

Examples of each type of sentence with both of the constructions and the English translations are given below in Table 15; a complete list of all experimental sentences used in the study is given in Appendix 2.

Table 15. Example stimuli used in sentence rating task

	Condition	Example	Length	WO	CX
1.	ade_short_lm1	Diivanil on ilus suur padi.	short	Lm1	ade
2.	peal_short_lm1	Diivani peal on ilus suur padi.	short	Lm1	peal
		'On the couch is a beautiful big pillow.'			
3.	ade_short_lm2	Ilus suur padi on diivanil.	short	Lm2	ade
4.	peal_short_lm2	Ilus suur padi on diivani peal.	short	Lm2	peal
		'The beautiful big pillow is on the couch.'			
5.	ade_long_lm1	Ilusal vanal diivanil on padi.	long	Lm1	ade
6.	peal_long_lm1	Ilusa vana diivani peal on padi.	long	Lm1	peal
		'On the beautiful old couch is a pillow.'			
7.	ade_long_lm2	Padi on ilusal vanal diivanil.	long	Lm2	ade
8.	peal_long_lm2	Padi on ilusa vana diiani peal.	long	Lm2	peal
		'The pillow is on the beautiful old couch.'			
9.	Control_Lm	See on ilus vana diivan.	control		
		'This is a beautiful old couch.'			
10.	Control_Tr	See on ilus suur padi.	control		
		'This is a beautiful big pillow.'			

Stimuli: experimental sentences. The experimental sentences were constructed by combining a "sentence stub" with a completion consisting of a Landmark phrase, a Trajector phrase and a verb. The Landmark phrase either followed the Trajector phrase, as in example 59, or preceded it, as in example 60:

(59) Trajector phrase Verb Landmark phrase

(60) Landmark phrase Verb Trajector phrase

The completion for the Landmark phrase consisted of either a one word noun phrase (examples 61 and 62) or a three word noun phrase – a noun modified by two adjectives (examples 63 and 64). With the adessive case construction, there is agreement and the adjectives are inflected for the adessive case as well (e.g. *ilusa**l** vana**l** diivani**l*** 'on the beautiful old couch' in example 63); with the adpositional construction, the adjectives agree with the noun and take the genitive case (e.g. *ilus**a** van**a** diivan**i** peal* 'on the beautiful old couch' in example 64).

(61) *diivanil*
couch.SG.AD
'on the couch'

(62) *diivani* *peal*
couch.SG.GEN on
'on the couch'

(63) *ilusal* *vanal* *diivanil*
beautiful.SG.AD old.SG.AD couch.SG.AD
'on the beautiful old couch'

(64) *ilusa* *vana* *diivani* *peal*
beautiful.SG.GEN old.SG.GEN couch.SG.GEN on.AD
'on the beautiful old couch'

The completion of the Trajector phrase also consisted of a one word noun phrase (e.g. *padi* 'pillow') or a three word noun phrase – a noun modified by two adjectives (e.g. *ilus suur padi* 'beautiful big pillow'). All experimental sentences contained the verb *olema* 'be' in the present tense, 3rd person singular form. Thus, all the experimental sentences were 5 words long and contained three elements.

Stimuli: Control sentences. The control sentences were constructed to control for any lexical effects and were of two types – control sentences with a long Landmark phrase (e.g. the row titled '9.' in Table 15 above) and control sentences with a long Trajector phrase (e.g. the row titled '10.' in Table 15 above). There were four sentences in each condition according to the four lexicalizations of Landmarks and four lexicalizations of Trajectors, giving a total of eight control sentences. In order to ensure that when participants rate the adessive and *peal*-sentences with long Landmark phrases, it is not the lexicalisation of the specific phrase itself that determines the difference in the ratings, but the construction, the ratings of the control sentences are compared. Ideally, control sentences should receive a similar rating. Should this be the

case, any differences detected between the sentences can be attributed to the variables length, word order, construction or a combination of them.

Constructing the questionnaire. In total, four Landmark and Trajector pairs were used to construct the experimental sentences: *table-vase, chair-bag, bed-blanket, couch-pillow*. The criteria for selecting these specific Landmarks was that they should be semantically similar, i.e. denote the same set of objects, in this case furniture, and that they should be used with sufficient frequency with both locative constructions. For each of the 4 Landmark and Trajector pairs eight sentences were constructed corresponding to the experimental conditions in Table 15, making the study a 2 (length: short vs. long) x 2 (word order: landmark first vs. landmark second) x 2 (locative construction: adessive vs. *peal*) cross-factorial design. Since there were four Landmark-Trajector pairs selected for each of the eight conditions (making it a total of 32 experimental sentences) and eight control sentences, there were 40 sentences in total in the questionnaire. The participants saw and rated all of the 40 sentences. There were two randomizations of the same set of stimuli. A full list of the sentences used in the rating task is given in the Appendix 2.

Participants. 103 native speakers of Estonian aged between 15 and 64 (mean age 33 years) participated in the experiment. 93 of the subjects were women and 10 men. 44 of the subjects came from South Estonia, 26 from North Estonia, 13 from Central Estonia, 11 from West Estonia and 9 from East Estonia.

Procedure. Participants were asked to complete a questionnaire distributed via the Internet using the eFormular tool and they were given the following instructions:

> The questionnaire is part of a study of Estonian locative cases and spatial adpositions. It takes about 5-7 minutes to complete the questionnaire. The questionnaire is anonymous. It is not an intelligence test or a grammar test; there are no right or wrong answers. I am only interested in the language use.
> Your task is to rate on a scale from 1 (not very probable) to 7 (very probable) how probable it is that you yourself would use the sentences given in the questionnaire. Read the sentences carefully, but do not spend too much time on it – I am interested in your initial reaction.

The instructions were followed by two examples for which ratings were provided. The first example sentence was given a rating of 1 and the second 7. These examples were provided in order to anchor the participants' ratings. The questionnaire was distributed electronically which enabled to ensure that the participants could not see their previous ratings and could not move back to change their responses. Participants were randomly assigned to one of the two randomizations, with 54 completing the first randomization and 49 the second.

3.3 The univariate statistical concepts and analysis techniques

The remaining two sections of this chapter on methodology take a look at the statistical techniques employed in the analysis of the corpus and experimental studies described above. The choice of adequate analysis techniques is not as straightforward as one might expect. Much depends on the nature of the data and the designs used to collect the data. Each specific statistical technique comes with its own set of assumptions which need to be checked before one can use it to analyse the data. One of the crucial factors in deciding which technique to use is the nature of the response and explanatory variable(s). As we saw above in the previous section, measures on a Likert-type scale can be approached from two different perspectives, i.e. treating the rating as interval (numeric) or ordinal (categorical). Following is a short discussion on some of these aspects. This section gives an overview of the univariate techniques employed in analysing the results that are presented in Chapter 4 and the next section focuses on some of the more complicated multivariate techniques. One of the reasons why I have included univariate analysis of the data in the thesis is that it allows the researcher to see the individual effects of each studied variable in isolation (Arppe 2008: 73). Nevertheless, one should never lose sight of the multivariate nature of language use and such univariate analyses as described below should be used only as the first steps leading to the more sophisticated multivariate techniques. Both descriptive statistics (summarizing the information in the data) as well as inferential statistics (providing predictions about a population, based on data from a sample) are made use of. The corpus and experimental studies make use of the following univariate methods and concepts: contingency tables, statistical hypotheses, significance level, chi-squared test, standardized Pearson residuals, *t*-test, Wilcoxon Mann-Whitney rank sum test, power, sample size, effect size, independent and dependent samples.

3.3.1 Chi-squared test

A common starting point in a univariate analysis of frequency data (i.e. the corpus data) is to represent it in **contingency tables** (also referred to as cross-classification or cross-tabulation). These tables allow to cross-tabulate the frequencies of the studied variables with each of the two constructions. Informative as such tables of raw frequency counts and percentages are, they do not allow us to draw any definitive conclusions as to whether the results are significance or just chance. Fortunately, we can use statistical techniques to assess and interpret the distributions represented in the tables. It is important to bear in mind that in the majority of the cases both the response variable (i.e. the locative construction) as well as the explanatory variables (i.e. the different semantic and morphosyntactic variables) in the corpus studies are nominal and

non-ordinal[47]. For nominal and non-ordinal data we want to know whether a statistically significant difference in distribution may simply have arisen due to the size of the data sample rather than from the strength of association between the variables (Arppe 2008: 75). We can assess the distribution of a variable with a statistical test of independence that the distribution in a contingency table deviates from the **null hypothesis** (H_0). The null hypothesis says that "nothing is happening" (Crawley 2005: 3). Specifically for the present thesis, the H_0 states that should the two locative constructions be synonymous, it would be expected that the observed frequencies of the levels of the different explanatory variables are evenly distributed between the constructions. If the null hypothesis holds, neither of the two variable types, i.e. the locative construction and the specific semantic or morphosyntactic variable, has an observable and statistically significant bearing on the other and the two variables are therefore independent of each other, in the statistical sense. If the null hypothesis does not hold, we have reason to reject H_0 and to assume that the **alternative hypothesis** (H_1) may be true – that the two variables are to some extent dependent on each other.

In practice what is evaluated for contingency tables is how strongly the observed frequencies deviate from the expected frequencies. One of the most commonly used measures in assessing the overall independence or dependence in a contingency table is the **Pearson chi-squared (X^2)**. The X^2 value together with the corresponding degrees of freedom is used to yield the **level of significance** according to the chi-squared distribution. The significance level (p-value or α) indicates the probability that the observed values in the contingency table could have been sampled by chance from the assumed population. In the humanities, the critical p-value is conventionally set at $p < .05$ and will be used also in this study. Statisticians warn, however, that rejecting the null hypothesis on the account of a significant p-value does not automatically lead to a direct confirmation of the alternative hypothesis (Cohen 2009: 4–6). If the p-value is below the critical value we can reject the null hypothesis of independence and assume instead that there is a strong association between the type of locative construction and the particular variable. The formula for calculating the X^2 consists of cell-by-cell calculations of the squared deviations of the observed values from the expected values. These cell-by-cell calculations are known as X^2 contributions and their square roots as Pearson residuals (Arppe 2008: 79).

Even if the chi-squared test by itself does tell us whether there is something significant overall in the relationship between the type of locative construction and the studied factor, it says very little about the exact locus or the direction of this association. We therefore need to look at to what extent individual cells account for the overall deviation from the expected values and in which

[47] According to Arppe (2008: 75 Ftn. 42) the *Stevens scale* that classifies data types into *nominal, ordinal, interval,* and *ratio* can be seen as a useful tool for linguists for selecting appropriate statistical methods. However, he also rightly points out that this scale "is not an exception-less absolute straightjacket" (Arppe 2008: 75 Ftn.42).

direction (either above or below) they deviate for these cell-wise contributions (Agresti 2002: 81). A convenient method to assess both the significance and direction of the cell-by-cell contributions is to calculate the **standardized Pearson residuals** (Agresti 2002: 81). Residual is the difference between an observed and expected cell frequency. In order to know whether a residual is large enough to indicate a departure from independence that is unlikely to be due to mere chance, a standardized form of the residual is used to obtain this information (Agresti and Finlay 2009: 230). Agresti and Finlay (2009: 230) state that "[t]he standardized residual is the number of standard errors that falls from the value of 0 that we expect when the H_0 is true". A large standardized residual provides evidence against independence in that cell, i.e. it points towards the conclusion that there is a significant association between variables and that the specific cell with a large standardized residual contributes to this association. Arppe (2008: 83) reports that in the 2x4 contingency table (a table he deems relatively small), a standardized Pearson residual which exceeds at least 2 in absolute value indicates a significant deviation in the cell in question; however, he also points out that no exact values have been provided in the literature. The present thesis follows the suggestions of Arppe (2008) and takes 2 as the benchmark for a significant deviation.

The basic **assumptions** of chi-squared test are that the variables are categorical and independent observations are randomly sampled (Agresti and Finlay 2009: 228). The extraction procedure described in section 3.1.1 above demonstrates that the data for the corpus study are randomly sampled and the use of two different corpora (i.e. two different populations) enables to treat the two samples (i.e. the sentences with the two constructions) as independent samples. A further assumption of the test is that the expected count in a cell of a contingency table is larger than 5 (Agresti and Finlay 2009: 228)[48]. As is seen in the next chapter, this assumption is also met.

3.3.2 T-test

In addition to the chi-squared test which is the appropriate test for drawing conclusions about the population on the basis of observations (e.g. frequency counts) from a sample where the measurements are nominal, the ***t*-test** has to be discussed, since the morphosyntactic variable 'length' in the multivariate corpus study is measured on a ratio-scale[49]. The *t*-test is frequently used to assess the differences between two population means. In choosing the appropriate *t*-test we have to distinguish between independent samples and dependent or paired samples. As we saw above in the discussion on chi-squared test, the multivariate corpus study constitutes an independent sample. The rating data,

[48] Fisher's Exact Test can be used for an analysis of contingency tables in which one or more of the expected frequencies is less than 5 (Crawley 2005: 90).

[49] As discussed above, the ratings data of the two experiments may be taken as measurements on the ratio-scale as well.

however, represents a dependent sample, since the ratings given for the two constructions can be paired up – each participant rated sentences with both types of constructions. The two different types of *t*-test are both relevant: the independent samples *t*-test and the dependent samples *t*-test. The **independent samples *t*-test** is used to test the null hypothesis that the means of two populations are the same; the test statistic is essentially a standardized difference of the two sample means (Everitt and Horton 2010: 49–50). The **dependent samples *t*-test** or paired *t*-test is used when the data come from paired observations, e.g. the two measurements are made on the same individual or were taken from the same location (Crawley 2005: 81). This is the case with the rating tasks, where each participant rated both constructions. The paired *t*-test is seen as more powerful or sensitive than the independent *t*-test, because paired observations from the same participant provide a control over individual sources of variation (Johnson 2008: 81).

The basic assumptions of *t*-test are that the data are obtained using randomization and the quantitative variable is assumed to have a normal population distribution (Agresti and Finlay 2009: 147). It is pointed out in statistical literature that as the sample size increases, the assumption of a normal population becomes less important due to the Central Limit Theorem (Agresti and Finlay 2009: 154–155). The Central Limit Theorem states that "for random sampling with a large sample size n, the sampling distribution of the sample mean $\bar{y}$ is approximately a normal distribution" (Agresti and Finlay 2009: 93). How large n must be before the sampling distribution is normal depends on the skewness of the population distribution; for most cases, a sample size of about 30 is sufficient (Agresti and Finlay 2009: 94). All the datasets in the present thesis exceed a sample size of 30. In case the assumptions of the *t*-test are not met, non-parametric [50] analogues are available. For two independent groups, the **Wilcoxon Mann-Whitney rank sum test** can be used and for paired samples, the **Wilcoxon signed rank test** is an available option (Everitt and Horton 2010: 51). The null hypothesis to be tested is that the two populations have identical distributions; the alternative hypothesis is that the population distributions differ in the median. Since there is controversy in considering the rating data as interval or ordinal, both the *t*-test as well as the rank test is used in analysing the data.

3.3.3 Power, sample size, and effect size

There has been criticism on the standard practice in statistical analysis of focusing the decision to reject or not to reject the null hypothesis on the basis of pre-selected *p* value (Cohen 1992, 1994). Rosnow and Rosenthal (1989: 1279), among others, stress that highly significant *p*-values should not be interpreted as

[50] Non-parametric statistical methods make no assumptions about the shape of the population distribution (Agresti and Finlay 2009: 209). Such methods are, however, less powerful in detecting differences.

automatically reflecting large effects. Instead, three other relevant statistical variables should be taken into account – statistical power, minimum sample size (*N*), and Effect Size (ES) – and the recommended procedure is to take into consideration all of these variables together with the significance level (α) (Arppe 2008: 80). As Cohen (1992: 156) emphasises, for any statistical model, the relationships between the four variables (power, sample size, effect size, and significance) are such that "each is a function of the other three". As we already established above, the significance criterion is set at $\alpha = .05$ in the humanities.

Cohen (1992: 156) defines statistical **power** as "the long-term probability, given the population ES, α, and *N* for rejecting H_0." A convention proposed for general use is to specify power at .80; a smaller value would lead to a too great risk of a Type II error[51] and a larger value would lead to a demand for *N* that is likely to exceed the researcher's resources (Cohen 1992: 156). By setting statistical power at .80 we are acknowledging that there is 80% chance of detecting an effect (i.e. rejecting the null-hypothesis).

A vital step in every research planning is to know the **sample size** necessary to obtain the desired power for the specified α and the hypothesised effect size. It is important to bear in mind that the sample size increases with an increase in the power, a decrease in the effect size, and a decrease in α (Cohen 1992: 156). The most difficult of the four relevant statistical variables to define is **effect size**, i.e. a measure of the strength of the association between two variables. Cohen (1992: 156) suggests that the reason may be that researchers do not fully comprehend the magnitude of the phenomena that characterizes much of psychological (or linguistic) research. However, the researcher does need to have at least some idea about the degree to which the H_0 is rejected (i.e. the effect size). Cohen (1992: 156) proposes small, medium, and large values for effect sizes across the different ES indexes, since each statistical test has its own effect size index; medium effect size represents an effect likely to be visible to the eye, small effect size is smaller than medium, but not trivial.

Cohen gives the effect size indexes with their specific values for small, medium, and large effects (1992: 157, Table 1) and specifies the minimum sample sizes for each effect size at power = .80 for $\alpha = .01$, .05, and .10 (1992: 158, Table 2). Both of these tables have been used in the thesis for the statistical analyses of the results. For contingency tables of the size studied in the present thesis, with the following different degrees of freedom[52] – df=1, df=2, df=3, df=4, df=6 – fixing α=.05 and *Power*=.80, Cohen (1992: 158, Table 2) has

[51] Type II error is the error of accepting the null hypothesis when it is false. Conversely, Type I error is rejecting the null hypothesis when, in fact, it is correct (Crawley 2005: 9).

[52] Crawley (2005: 37) proposes the following formal definition of degrees of freedom: degrees of freedom is the sample size minus the number of parameters estimated from the data. The degrees of freedom for contingency tables are defined as $df = (r - 1) \times (c - 1)$, where *r* is number of rows and *c* number of columns. Therefore, for a 2x2 contingency table we have $(2 - 1) \times (2 - 1) = 1$ degrees of freedom. (Crawley 2005: 88). The important point here is that knowing the values of say five numbers, we have total freedom in selecting the first four numbers, but no choice at all in selecting the fifth number.

calculated the minimum sample sizes[53] given in Table 16. This means that, for example, for a contingency table with 1 degrees of freedom, 26 observations are needed to detect a large effect size, 87 for a medium effect, and 785 for a small, but not trivial, effect.

Table 16. *N* for Small, Medium, and Large effect size at Power = .80 for α=.05 for chi-squared test (extracted from Cohen 1992: 158, Table 2)

X^2 test	Small ES	Medium ES	Large ES
1df	785	87	26
2df	964	107	39
3df	1,090	121	44
4df	1,194	133	48
6df	1,362	151	54

Given that the multivariate corpus study is based on 450 sentences for each construction, the sample size is sufficient for detecting medium and large effects. Although detecting small effects would ideally be desirable as these effects are not trivial, the nature of linguistic data is such that the focus should be instead on larger effects. Language is never random and the potential of finding small effects of any two variables is relatively high. The effect size index for the chi-squared test of independence is referred to as ***w***. The values of the effect size index *w* for small, medium and large effect size are **.10**, **.30**, and **.50** respectively (based on Cohen 1992; 158, Table 2). In case of 2x2 contingency tables or for any table with either two rows or two columns (as are all of the contingency tables reported in the present study) the effect size is equal to ***Cramér's V*** (see Arppe 2008: 86–94 for an overview and discussion of different measures of association[54]).

According to Cohen (1992: 156–157) the effect size index for the *t*-test of the difference between independent means is ***d*** – the difference expressed in units of the within-population standard deviation. The values of the effect size index *d* for small, medium and large effect size are **.20**, **.50**, and **.80** respectively. Table 17 gives the minimum sample sizes for calculating difference between two independent sample means at α=.05 and *Power*=.80 (Cohen 1992: 158, Table 2). This means that 26 observations are needed to detect a large effect size, 64 for a medium effect, and 393 for a small effect. Both the corpus observations as well as the rating task data are large enough to detect even a small effect for the difference between the means of the two constructions.

[53] For statistical tests involving two groups (as the two construction types in the present studies), *N* in Table 16 is the necessary sample size for each group (Cohen 1992: 156).

[54] As Arppe (2008: 86) points out, in case of nominal variables the relationship between the studied variables is referred to by the term *association*, instead of *correlation* which is reserved for interval variables.

Table 17. *N* for Small, Medium, and Large effect size at Power = .80 for α=.05 for *t*-test of two independent samples (extracted from Cohen 1992: 158, Table 2)

t-test	Small ES	Medium ES	Large ES
mean difference	393	64	26

In discussing the issue of power in acceptability judgement experiments, Sprouse and Almeida (submitted) report that a sample size of 11 participants for a forced-choice task, each providing only a single judgement per condition, and 35 for a magnitude estimation task, is sufficient for 80% power, which as we saw above, is considered the target power level for experiments in psychology and the related fields. It is also pointed out that increasing the number of judgements per condition increases statistical power, thereby decreasing the required sample size (Sprouse and Almeida submitted). The conclusion they draw (*ibid*.) is that a non-numeric task such as a forced-choice decision on acceptable vs. unacceptable sentences tends to be more powerful than numerical rating tasks at detecting differences. However, as argued above, non-numerical tasks do not say anything about the size of the difference. In addition to emphasising the need to estimate the required sample size for a given research phenomena before collecting the data, Schütze and Sprouse (to appear) also remind us that the failure to find an effect in the data does not mean that there is no effect; it may simply mean that the sampling procedure (e.g. the experimental design) is not powerful enough to detect the difference (Schütze and Sprouse to appear: 16).

3.4 The multivariate statistical techniques

Univariate analysis helps the researcher to identify which variables are statistically relevant and linguistically meaningful and it is a necessary stage before moving on to multivariate analysis (Arppe 2008: 71). Multivariate analysis takes into account the fact that the variables have a joint and simultaneous influence on the linguistic phenomenon under study and it allows the researcher to assess the relative weight of specific variables in relation to each other (Arppe 2008: 113). By employing multivariate statistical techniques to analyse the data, the underlying assumption is that the alternation between the two constructions is influenced by a number of variables simultaneously. There are numerous multivariate techniques available (see, for example, the lists provided in Rencher 2002 or Everitt 2005) and the choice, again, depends on a number of issues. One of the deciding factors in choosing a technique is the nature of the response and explanatory variable(s). Relevant for the present study are multivariate techniques that apply to datasets where the dependent variable is categorical (either nominal, more specifically binary, as in the multivariate corpus study or ordinal as in the two rating tasks) as well as

numeric (if we treat the rating in the experimental tasks to be a measurement on an interval scale). Explanatory variables are also of two types – categorical (as the majority of the explanatory variables in the corpus studies as well as all the explanatory variables in the rating tasks) or numeric (e.g. length of phrases as measured in the number of words or syllables in the multivariate corpus study). The present study employs the following multivariate techniques: binary logistic regression, ordinal logistic regression, and analysis of variance. All three techniques involve regression analysis (Faraway 2002, Rencher 2002, Baayen 2008: 165–238, Dalgaard 2008: 227–248, Everitt and Horton 2010: 117–136), which in addition to indicating which variables influence the response variable most, allow access to assessing how well these results are able to predict future observations. Regression analysis is used for explaining or modelling the relationship between a single variable, called the *response, output* or *dependent* variable, and one or more *predictor, input, independent* or *explanatory* variables. Depending on the number of parameters[55], it can be called a simple regression ($p = 1$) or a multiple regression ($p > 1$) (Faraway 2002: 13).

All three multivariate techniques discussed in the thesis (binary logistic regression, ordinal logistic regression, and analysis of variance) are, in essence, quite similar and can be unified under the name of *generalised linear model* (GLM) (Everitt and Horton 2010: 121). All three involve a linear combination of a set of explanatory variables as a model for the observed response variable (Everitt and Horton 2010: 121). However, with logistic regression models (binary as well as ordinal), the expected value of the response is not modelled directly but via a logistic transformation. The usefulness of generalised linear models resides in their powerful and flexible nature – these models can be used when the variance is not constant, and/or errors are not normally distributed (Crawley 2005: 113). The structure of the model relates each observed value to a predicted value which is obtained by transformation of the value emerging from the linear predictor; the latter is sum of the effects of one or more explanatory variables (Crawley 2005: 115).

Depending on the nature of the effects of the variables, a further distinction is made between **fixed-effects models** and **random-effects models**; models that take both kinds of effects into account are known as **mixed-effects models**. The present study makes use of fixed-effects models (for the corpus data) and mixed-effects models (for analyzing the rating tasks). Fixed effects are parameters associated with an entire population or with repeatable levels of experimental factors (Pinheiro and Bates 2000: 3). They are unknown constants to be estimated from the data (Crawley 2005: 178) and I am interested in the specific levels of these variables. As such, we can treat the various semantic and morphosyntactic explanatory variables in the corpus study and the experimental treatments (i.e. the variables of construction type, Landmark type, length of the Landmark phrase, and the relative word order between the Landmark and Trajector) in the two rating tasks as fixed effects. Random effects, on the other

[55] A parameter is a summary measure of the population (Agresti and Finlay 2009: 5).

hand, are associated with individual experimental units drawn at random from a population (Pinheiro and Bates 2000: 3). With random effects, we are interested in how much of the variance in the data they explain (Crawley 2005: 178), but not in the specific levels of these effects. As such, we can treat the specific items and participants in the two rating tasks as random effects. We can say that they are sampled randomly from populations of words and participants, and replicating the experiment would involve selecting other words and other participants (Baayen 2008: 241). If we do not take subjects and items into account as random variables, the model does not generalize to the population of subjects and items, i.e. the results are applicable only to the specific levels/types of subjects and items in these two rating tasks. Statistical literature makes a crucial distinction between fixed-effects terms and random-effects terms, but it is not always clear to decide when to use a specific explanatory variable as a fixed effect and when as a random effect (cf. Crawley 2005: 179–180 for some general guidelines). To highlight the necessity of taking subjects and items into account as random effects in analysing the results of the experiments, Chapter 5 presents both types of models – models that only include fixed effects and models that include both fixed and random effects. It will be shown that the mixed-effects models do a better job at providing a more adequate and powerful description of the data.

All in all, the many benefits of regression analyses include the following: prediction of future observations; assessment of the effect of, or relationship between, explanatory variables on the response; and the description of the data structure in general (Faraway 2002: 13). Fitting a regression model used to require extensive hand calculations, but with high-speed computing, regression methodology developed rapidly and the scope for analysis has widened (Faraway 2002: 15). The analysis of the data discussed in the present thesis is carried out using R (version 2.10.1, R development core team 2009). R is an open-source language and environment for statistical computing, freely available at http://www.r-project.org. The subsequent sections give a short overview of all these techniques; a discussion on some of the general issues, e.g. model selection and assessment, relevant for all of these multivariate techniques precedes these sections.

3.4.1 Model building and model checking

The main aim of statistical modelling is to determine the values of the parameters in a specific model that lead to the best fit of the model to the data (Crawley 2005: 4). The best model is the model that produces the least unexplained variation (minimal residual deviance), all the parameters in the model are statistically significant, and following the principle of parsimony, it is minimally adequate (Crawley 2005: 4). It is important to realise that there is no one sacred model – a number of different models can adequately describe the data. It is the job of the researcher to determine which of the possible models is the "best" model, i.e. **the minimal adequate model**. The 'best fit of the model

to the data' is generally defined in terms of maximum likelihood – we evaluate the model on the basis of how likely the data would be if the model was correct (Crawley 2005: 5–7). Crawley (2005: 7), like many other statisticians, places a special emphasis on model simplification – the principle of parsimony or **Occam's razor**. This principle states that given a set of equally good explanations for a given phenomenon, the correct explanation is the simplest explanation (i.e. explanations are "shaved" down to the bare minimum). For statistical modelling this means that the models should have as few parameters as possible and simple explanations should be preferred to complex ones, at the same time bearing in mind Einstein's words that 'a model should be as simple as possible. But not simpler.' (cited in Crawley 2005: 8). The following truths have to be remembered about any model: "all models are wrong, some models are better than others, the correct model can never be known with certainty, and the simpler the model, the better it is" (Crawley 2005: 119).

Statistical handbooks describe a number of model selection and model checking procedures (e.g. McCullagh and Nelder 1989: 21–43, Dobson and Barnett 2008: 45–55, Agresti and Finlay 2009: 441–474). A crucial step in model building involves **variable selection**, i.e. specifying which explanatory variables enter the model. For example, the multivariate corpus study looks at 20 variables in total, but the minimal adequate model should only include the variables that are significant. Unnecessary variables add noise to the estimation of the predictors we are interested in, and collinearity (see below) is often caused by having too many variables trying to do the same job (Faraway 2002: 124). According to Faraway (2002: 125) the simplest of all variable selection procedures is **backward elimination**, which starts with all the predictors in the model and proceeds by the one-by-one removal of factors that have a higher *p*-value than the critical α-level (α_{crit}). Faraway (2002: 125) notes that the α_{crit} does not have to be 5% and a 15-10% cut-off may work as well; for the present study the conventional 5% cut-off is selected. The reverse procedure of variable selection is known as **forward selection** (Faraway 2002: 125–126). **Stepwise regression** is a combination of backward elimination and forward selection, where variables may be added or removed at each stage (Faraway 2002: 126–128). For the present study, stepwise regression is used alongside criterion-based procedures for the variable selection procedure. Faraway (2002: 126) stresses that with any variable selection method, it is important to keep in mind that we cannot separate model selection from the purpose of the study. The statistical significance of the variables that stay in the model tends to be amplified and it would be wrong to claim that the variables that are dropped are not correlated with the response variable – they just do not add enough explanatory effect over those variables that are already included in the model for the model to become significantly better (Faraway 2002: 126).

Besides stepwise procedures, there are also **criterion-based procedures** available for model selection. The following three are some of the more commonly used criteria (these are also used in the present study): the Akaike Information Criterion (AIC), the Adjusted R^2 known as R_a^2, and C-score. **AIC** is

used for finding the simplest model and it penalizes the model for having more parameters than are useful for getting good predictions (Agresti and Finlay 2009: 448). The aim is to minimize the value of AIC and the "best" model is the one with the smallest AIC. $\boldsymbol{R_a^2}$ [56] is used as a measure for comparing the predictive power of different models. R_a^2 represents the proportion of variance explained by a specific model and the larger the value of this measure is, the better. Another important score is the **C-score**, which is an index of the correlation between predicted probability of expected responses and actual responses. As with R_a^2, the larger the value of C-score, the better. The advantage of criterion-based procedures over stepwise procedures is that the former do not use the dubious *p*-values (cf. Cohen 1994 for a critical overview of what is wrong with null hypothesis significance testing and making mechanical decisions around a sacred .05 criterion). Like many other statistical techniques and procedures, variable selection is a means to an end and not the end itself and the described methods should be used as guides with the relevant cautionary notes in mind. Both stepwise procedures as well as criterion-based methods are used in comparing the models in this thesis. It is important to remember that there always is a trade-off between model simplicity and fit; the ideal model is a compromise between the two (Crawley 2005: 208).

Another aspect of model building involves the **diagnostics** – after selecting the predictors for a model, the next step is finding out if the model fits the data adequately. Two issues are at stake here: being able to identify whether the assumptions of the model are violated, and finding if certain observations are highly influential in the model fit. As is mentioned above, it is virtually impossible to verify that any given model is 100% correct and model diagnostics mainly involves checking that the model is not grossly wrong (Faraway 2002: 14). Only if the chosen model describes the data adequately, conclusions can be drawn about the parameter estimates, i.e. about the weight and type of impact of explanatory variables on the response variable. Various diagnostics are used to detect problems with a model, to suggest improvements and to ascertain whether it provides a satisfactory fit to the data (Faraway 2002: 72).

Faraway (2002: 14) points out that graphical diagnostic methods are generally preferred over numerical diagnostics, because the former are more versatile and informative. In the present thesis, the R function `plot()` is made use of for graphical diagnostics. One of the basic diagnostic quantities used are **residuals** (Faraway 2002: 72, Baayen 2008: 188). Residuals are the differences between the data, i.e. the measure value, and the fitted model, i.e. the value predicted by the model (Crawley 2007: 389). The residuals of the models

[56] R^2 by itself is not a good criterion because it would always choose the largest possible model (Faraway 2002: 129). That means that if we take this measure as the criterion for selecting the "best" model, we will end up with the most complicated model because it has the highest value for R^2; in such a case we will be violating the principle of parsimony (Agresti and Finlay 2009: 446).

described in the present thesis have been checked by plotting the residuals against the fitted values to assess model fit. Another aspect about running the model diagnostics involves identifying the **outliers** – data points that do not fit the model (Faraway 2002: 75). Outliers can be data entry errors, truly unusual points, or residuals which are large but not exceptional. Ideally, outliers of the first sort should be corrected and outliers of the second sort reported, but excluded from the model in order to improve model fit (Faraway 2002: 77). As Baayen (2008: 202) points out, "the distorting presence of just a few atypical outliers may obscure effects that characterize the majority of data points". Residual plots are not only used for identifying outliers and influential data points, but also for checking the assumptions of the model. By plotting the residuals against the fitted values we can detect heteroscedascity (non-constant variance) and nonlinearity (which indicates that it is necessary to make changes to the model) (Faraway 2002: 80–87).

Standard procedure for regression modelling involves comparing the fit of competing models – models with different subsets of explanatory variables. A measure of fit in regression analysis is based on deviance which measures how closely the model-based fitted values of the response variable approximate the observed values and any two models can be compared by comparing the respective deviance values (Everitt and Horton 2010: 122). Changes in deviance caused by a model reduction will be approximately χ^2-distributed with degrees of freedom equal to the change in the number of parameters (Dalgaard 2008: 228). This procedure is also used in the present thesis to compare the fit of competing models.

It is common in social science studies (and the related fields like linguistics) that use multiple regression that the explanatory variables "overlap" considerably. This condition is referred to as **multicollinearity** in the statistical literature. In case of multicollinearity, explanatory variables enter into strong correlations and it becomes very difficult to tease apart the explanatory values of the correlated variables (Baayen 2008: 181). Multicollinearity inflates standard errors for estimates of regression parameters (Agresti and Finlay 2009: 456). As we saw above in section 2.3, several operationalizations of the same variable are included in the multivariate corpus study. It was not entirely clear which measure would capture the essence of the factor the best (e.g. measuring length in words vs. in syllables) and it was thought best to include both. Other instances of collinearity include, for example, the type, mobility, and animacy of Trajectors and Landmarks – although there is no one-to-one correspondence between the different levels of these factors, it is clear that these factors are strongly correlated. An animate Trajector or Landmark is categorised under the level of 'things' in the variable 'type' and they are also mobile. It is therefore necessary to model these variables separately and decide on the basis of comparing different models (according to the measures described above) which models provide the best fit and have the most predictive power. Most statistical software display the **variance inflation factor (VIF)** to describe the extent to

which multicollinearity exists. This measure is also used in the present study to check the existence of multicollinearity.

The issue of multicollinearity ties in with the issue of epiphenomenal variables in various alternation studies discussed in Chapter 2. For instance, in the alternation studies on the English genitive, dative and particle placement, it has been noted that variables like animacy, topicality, and weight correlate highly with each other (e.g. Hawkins 1994; Rosenbach 2005, 2008). Rosenbach (2005, 2008), for example, discusses whether animacy effects are simply an artefact of syntactic weight, or whether animacy is an independent variable in grammatical variation. She concludes that animacy is not an epiphenomenon of other variables in grammatical variation, despite its statistical correlation with variables such as topicality and weight. She makes reference to the following studies as converging evidence: Rosenbach (2002, 2005), Bresnan et al. (2007). Complexity and length is another pair of variables that correlate considerably with each other – longer constituents tend to be more complex and vice versa (Rosenbach 2005: 617). Wasow and Arnold (2003: 120–128) provide a nice overview of the controversy regarding whether length and complexity are distinct variables or not. These issues pose a problem for comparing the effects of two correlating factors and the answer, in my opinion, is partly provided by logistic regression. This multivariate method enables to assess the degree to which collinearity is a problem for a given model and it allows us to compare different models that include different explanatory variables, allowing the researcher to conclude which variables are more relevant for a specific alternation phenomenon.

3.4.2 Repeated measures ANOVA

According to Johnson (2008: 104), analysis of variance (ANOVA) is the gold standard approach in psycholinguistic experiments if participants provide responses that can be scored on continuous scale – such as reaction time, preference, or accuracy. It is the multivariate technique used when all the explanatory variables are categorical and the response variable is numeric. The "variance" part in the name of the method comes from the original idea of trying to partition the overall variance in the response to that due to each of the factors and the error (Faraway 2002: 168). In ANOVA, the explanatory variable is called a factor and each factor has two or more levels; the parameters are often called effects. Depending on the number of explanatory variables, distinction is made between one-way, two-way or three-way ANOVA. When there is replication at each level in a multi-way ANOVA (as in the two rating tasks reported here), the experiment is called a factorial design and it allows to study interactions between variables. In the two rating tasks, I am interested in finding out whether the response (or the rating given) to one factor (e.g. the locative construction) depends on the level of another factor (e.g. type of Landmark, length of the Landmark phrase, or word order). The type of ANOVA used to analyse the experimental results in the thesis is **repeated measures factorial ANOVA**: "repeated measures" because I have repeated measurements, i.e. ratings from one

and the same participant, and "factorial" because I am interested in the interaction between different factors. According to Everitt and Horton (2010: 82):

> "The model used in the analysis of variance leads to a partition of the variation in the observations into parts due to main effects and interaction plus an error term that enables a series of *F*-tests to be calculated that can be used to test hypotheses about the main effects and the interaction."

The general assumptions of the standard ANOVA are the following: scores on an interval scale, the observations are independent of each other, the observations arise from a population having a normal distribution and the same variance (Johnson 2008: 104–106, Everitt and Horton 2010: 83). The experimental designs reported here violate the assumption of independence. This is where the classical ANOVA has its limitations – it requires that the researcher test many more subjects than are actually necessary and it keeps from using each subject as his/her own control (Johnson 2008: 121). Using the repeated measures ANOVA allows to take into account the random unexplained variation contributed by the individual differences between the raters. As we will see in the next chapter, some participants give overall higher ratings than other participants. The observations from one participant are often more highly correlated with each other than they are with the observations from other participants (Johnson 2008: 122). At the same time, given that Participant 1 tends to give all of the sentences relatively high ratings, it would be particularly interesting to see if his or her ratings differ across experimental conditions, i.e. despite the overall difference in the ratings of Participant 1 an experimental manipulation may in fact impact the behaviour in a consistent way as well. Using a repeated measures design can therefore help to improve the precision of estimation (Agresti and Finlay 2009: 393).

A final important point about analysis of variance concerns the issue of effect size. Bakeman (2005: 379) acknowledges that there is some confusion as to which specific effect size statistic should be computed and reported in case of repeated measures ANOVA. He himself (Bakeman 2005) endorses the use of generalized eta squared (η^2_G) as the appropriate statistic for repeated measures designs, but Pierce, Block and Aguinis (2004) note that the accurate measure is **the partial eta-squared ($\boldsymbol{\eta^2_P}$)**. Both of these statistics differ from the classical *eta squared* (η^2), the effect size statistic for standard ANOVA which represents the ratio of effect to total variance (Bakeman 2005: 379). The difference is that for repeated measures ANOVA, the error term has to be taken into account as well when computing the effect size. Although Bakeman (2005: 380) stresses that if the aim is to provide comparability across studies, generalized eta squared clearly seems the best choice for an effect size statistic, partial eta-squared is reported in the present study since it is easier to calculate and seems to be the more common statistic reported in studies using repeated measures

ANOVA[57]. According to Pierce *et al.* (2004: 918), "partial eta-squared is defined as the proportion of total variation attributable to the factor, partialing out (excluding) other factors from the total non-error variation". Similarly to the classical eta-squared, the values for partial eta-squared also range from 0 to 1. From the information reported in the ANOVA summary tables in Chapter 5, partial eta-squared can be computed in the following way: partial $\eta^2 = SS_{factor}/(SS_{factor} + SS_{error})$[58], where SS_{factor} is the variation attributable to the factor and SS_{error} is the error variation (Pierce *et al.* 2004: 918). In the two rating tasks, participants account for the error variation. According to Cohen (1992: 157, Table 1), for analysis of variance, the values for small, medium, and large effect size indexes are **.10**, **.30**, and **.50** respectively.

3.4.3 Binary logistic regression

Many statistical problems involve binary response variables and the present research questions falls into this category as well – I am interested in the choice between two locative constructions, the adessive and the *peal*-construction. One possible technique to be used in such a case is binary logistic regression. A number of studies on grammatical alternations and lexical synonymy reviewed in Chapter 2 (e.g. Bresnan et al. 2007, Hinricsh and Szmrecsanyi 2007, Arppe 2008, Szmrecsanyi and Hinricsh 2008, Bresnan and Ford 2010, Shih et al. to appear, Szmrecsanyi in press b, Wolk et al. to appear) have also found (binary) logistic regression a highly valuable tool. In the simplest case of binary regression the response variable y takes only two values (Tutz 2012: 29). It may be coded by $y = 1$ and $y = 0$, where often $y = 1$ is considered as success and $y = 0$ as failure. A measure often reported by statistical software for categorical data is ***odds ratio***. The estimated ***odds*** for a binary response equal the number of successes divided by the number of failures (Agresti and Finlay 2009: 236). For example, if the probability of success is 0.75, the probability of failure equals $1 – 0.75 = 0.25$, and the odds of success equals $0.75/0.25 = 3.0$. When odds are 3.0, a success is three times as likely as a failure and we expect about three successes for every failure (Agresti and Finlay 2009: 235).

A binary response variable is distinctly different from a continuous response variable and certain transformations are required for doing regression analysis. Particularly important for logistic regression models are odds and log-odds or **logits** (Tutz 2012: 31). Statistical software use natural logarithms in fitting the model and rather than measuring association between two groups by odds ratios, one can use the log-transformed odds ratios (Tutz 2012: 33). The logit

[57] Since I am no expert on this issue, I will rely on the argumentation provided by Pierce *et al.* (2004) and report the partial eta squared in the ANOVA summary tables in Chapter 4. Readers should bear in mind, however, that this specific statistical conundrum, as well as many others, should not be taken as black and white truths, and that statistics like linguistics is a science riddled with controversies and unresolved issues.

[58] *SS* in statistics is used as an abbreviation for 'Sum of Squares'.

model is the most widely used model in binary regression (Tutz 2012: 37). In binary logistic regression, there is no error term or variance parameters – we are modelling the probability of an event directly and that in itself determines the variability of the binary outcome (Dalgaard 2008: 228). This technique involves the same model building and model checking steps as described above; these are applied for the present study as well.

3.4.4 Ordinal logistic regression

The appropriate multivariate technique for response variables that are categorical variables with ordered levels is ordinal logistic regression. Both rating tasks have response variables ordered on a scale from 1 to 5 and 1 to 7 respectively. With ordinal variables, it is possible to rank them, but the real distance between the categories is unknown. Two possible alternatives exist: it is possible to disregard the ordering of the values and treat the variable as if it was nominal, or you can treat the variable as it was interval (Norušis 2011: 69). While the first option clearly loses the information about the ranking, the second option is an attractive (although controversial) alternative and will be used in this study as well (c.f. analysis of variance). It would be extremely useful, however, if it were possible to apply a model that incorporates the ordinal nature of the response variable. Fortunately, there are statistical procedures available that allow this – the SPSS Ordinal Regression procedure or PLUM (Polytomous Universal Model) (Norušis 2011: 69) and the package `ordinal` in R (Christensen 2011); both of these are used in analysing the results of the two rating tasks. It is also possible to extend the binary logistic regression model to multinominal regression model, but for ordinal categorical variables the drawback of the multinominal regression model is that the ordering of the categories is ignored. Instead, it is possible to:

> "modify the binary logistic regression model to incorporate the ordinal nature of a dependent variable by defining the probabilities differently. Instead of considering the probability of an individual event, you consider the probability of that event and all events that are ordered before it." (Norušis 2011: 70).

In practical terms, the extension of a logistic regression model involves selecting a different link function[59] – since the ordinal logistic model is not linear it is necessary to do a transformation on the response variable. There are different functions available for ordinal regression procedures both in SPSS and R. The most typical function for ordinal logistic regression as well as for binary logistic regression is that of logit link function – the natural logarithm of the odds. The logit link function is applied when the categories are evenly distributed and the changes in the cumulative probabilities are gradual (Norušis

[59] "The link function is the function of the probabilities that results in a linear model in the parameters. It defines what goes on the left side of the equation." (Norušis 2011: 83)

2011: 85). Complementary log-log link function is available as an alternative in SPSS for datasets where the higher categories are more probable.

As with any regression model, the model needs to be checked for goodness of fit and other relevant issues. A standard model checking technique that assesses the model fit for ordinal logistic regression is to compare the observed and expected values (Norušis 2011: 75). This allows computing the Pearson and Deviance measures which provide further information about how well the model fits the data. If the model fits well, the observed and expected counts are similar, and the value for the Pearson and Deviance measures is small and the significance level is large (Norušis 2011: 78). It is also important to check the assumption of parallel lines, i.e. that the relationships between the explanatory variables and the logits are the same for all the logits (Norušis 2011: 74)[60]. If the assumption of parallelism is rejected, multinominal regression should be used as an alternative. The model -2 log-likelihood and chi-squared tests are used to assess the contribution variables make to the model. The pseudo *R*-square statistics can be used to measure the strength of the association between the response and explanatory variables (Norušis 2011: 81).

Although the standard ordinal logistic regression model is very useful, a further extension needs to be added. As we saw in the above discussion on mixed-effect models, it would be highly desirable if the random effects of participants and items could be added to the ordinal logistic regression model. This would enable to generalise the results to other participants and items as well. Unfortunately, the current SPSS Ordinal Regression procedure does not allow that, but the `clmm` function from the R package `ordinal` does allow it. Since this is a very recent development, I am aware of only one linguistic study that uses this specific statistical technique – Divjak (in progress)[61]. The current `clmm` function in R only allows one random term to be included in the model. Although R has the advantage over SPSS for allowing mixed-effects ordinal regression modelling, the advantage of SPSS over R is that it comes with already built-in model checking techniques. In R, the basic assumption of parallel lines has to be checked manually, as no procedure is implemented in the current version of the R package `ordinal`.

3.5 Summary of methodology and statistical techniques

This chapter outlines the designs of the two corpus studies and the two rating tasks. The above sections on the univariate and multivariate statistical techniques stress the importance of selecting the right technique for the analysis of any given dataset, taking into account the specific designs of each study. The selection of the right technique depends on the nature of the dataset – the nature

[60] The null hypothesis states that the slope coefficients in the model are the same across response categories; hence, lines of the same slope are said to be parallel.

[61] The analysis procedures and model checking described here is based on Divjak's (in progress) paper since no textbook guidelines exist.

of the response and explanatory variables and whether the data meet the assumptions posited by the specific techniques, e.g. normal distribution of errors, homogeneity of variance, independence of samples, etc. Another important aspect about using statistics is not to get too attached to the sacred *p*-value. Sometimes there may be a significant association between two variables, but it may very well be that the effect of this association is relatively small. The notions of 'power', 'sample size', 'effect size', and 'significance level' are all very important (not just *p*-values) and should be taken into account when undertaking statistical significance testing. With regression analysis it is especially important to consider other aspects than just including the statistically significant variables in the model – model checking and model criticism involve the overall assessment of how well a given model fits the data. No single model is 100% correct and there is no one single model that fits a given dataset. A researcher has to choose the "best model" from among a set of models, i.e. the minimally adequate model.

Table 18 summaries the design of the four studies reported in Chapters 4 and 5. It specifies in the second column the response and explanatory variables used in each study. The third column indicates which of the above-described statistical techniques are used to analyse the results of the four studies. It is important to note that this is only one possible selection and implementation of the valid techniques. Given the details about the designs and the nature of the datasets, it is decided that these specific techniques are the best ones to be used for the analysis; other researchers may opt for different techniques.

Table 18. Overview of the corpus and experimental studies

Study	**Response and explanatory variables**	**Statistical technique**
Multivariate corpus study	Response: categorical, binary (adessive vs. *peal*-construction) Explanatory: categorical and interval (11 semantic and 9 morphosyntactic variables)	chi-squared test *t*-test for independent samples binary logistic regression
Distinctive Collexeme Analysis	Response: categorical (frequency) Explanatory: categorical (type of Landmark)	Fisher exact test
Picture rating task	Response: interval, ordinal (rating on a scale from 1 to 5) Explanatory: categorical (type of Landmark & word order)	*t*-test for dependent samples repeated measures ANOVA
Sentence rating task	Response: interval, ordinal (rating on a scale from 1 to 7) Explanatory: categorical (length of the Landmark phrase & word order)	linear mixed-effects regression ordinal logistic regression ordinal mixed-effects logistic regression

4. RESULTS: CORPUS STUDIES

The previous chapter described the methodologies and statistical techniques used for analysing the results of the corpus and experimental studies reported in the thesis. The following two chapters present the actual results of these studies. This chapter reports the results of the two corpus studies. Sections 4.1 and 4.2 present the univariate and multivariate results of the first corpus study – the multivariate corpus study on the two locative constructions in Estonian, the adessive case and the adposition *peal* 'on'. The aim of the multivariate corpus study is to examine the effect of all the linguistic variables. The results are taken as the basis for formulating the hypotheses tested with the two rating tasks reported in Chapter 5. First, each explanatory variable is looked at individually, determining whether there is a statistically significant association between the specific semantic or morphosyntactic variables and the locative construction. Second, all the explanatory variables are looked at simultaneously and binary logistic regression is applied to analyse the results. Section 4.4 reports the results for the distinctive collexeme analysis, followed by a summary of the corpus results given in section 4.5. The analysis of the data is carried out using R (version 2.10.1, R development core team 2009).

4.1 Univariate results of Multivariate Corpus Study

As a reminder, the multivariate corpus study included eleven semantic variables and nine morphosyntactic variables (see Table 11 on p. 109) and looked at 900 corpus extractions. 450 adessive constructions were randomly extracted from fiction and newspaper subcorpora of the Morphologically Disambiguated Corpus of Estonian (size 215,000 words). 450 *peal*-constructions were randomly extracted from fiction and newspaper subcorpora of the Balanced Corpus of Estonian (size 10 million words). The reader is referred back to section 3.1.1 for details about the specific corpora and extraction procedures.

The results are presented in two broad categories – semantic and morphosyntactic variables, following the order used in Chapter 2. The first part of section 4.1 gives an overview of the univariate results for the following eleven semantic variables: type of relation between Trajector and Landmark, type, animacy, number, and mobility of Trajector and Landmark, relative size of the Trajector in relation to the Landmark, and verb lemma. The second part reports the univariate results for the nine morphosyntactic variables: length of Landmark, morphological complexity of Landmark, syntactic function of the Landmark phrase, word class of Trajector and Landmark, clause type, position of the Landmark phrase, and the relative position between the Trajector and Landmark phrase.

The strength and significance of the association between a specific explanatory variable and the response variable, i.e. the locative construction, is assessed using either the chi-squared test (for categorical explanatory variables)

or the independent *t*-test (for continuous explanatory variables) (see section 3.3 for overviews of these methods). For each explanatory variable, either a table or a graph is given, where the actual frequency counts or average measurements are presented. The tables presenting count data include not only the observed counts, but also expected counts (these are given in the brackets). This is followed by reporting the results of the statistical tests. The general conventions of reporting the results are followed – first the exact value of the relevant test statistic (either χ^2 or *t*) is given (with the degrees of freedom and sample size in the brackets), followed by the *p*-value and the effect size. The Pearson residuals for significant variables are given in a separate table – these play an important role in interpreting the strength and direction of the association between the variables. The paragraphs contain an initial discussion of the results, but the discussion is left for Chapter 6, where the results of the different studies presented in the thesis are compared to each other as well as to other relevant studies on alternations.

4.1.1 Type of relation between Landmark and Trajector

The first semantic variable – type of relation between Landmark and Trajector – has two levels: 'abstract' and 'spatial'. It is predicted that the case construction is the preferred construction when the relation is abstract. The frequency counts for this variable are given in Table 19. From Table 19 it can be seen that abstract relations are fairly infrequent (141 out of 900) compared to spatial relations (759 out of 900). Out of the 141 instances of an abstract relation between the Trajector and Landmark, the adposition *peal* 'on' is slightly more frequent (83) compared to the adessive (58).

Table 19. Type of relation between Landmark and Trajector and construction (expected frequencies in parentheses)

Relation type	adessive	peal	Total
abstract	58 (70.5)	83 (70.5)	141
spatial	392 (379.5)	367 (379.5)	759
Total	450	450	900

The results show that there is a significant although very small difference – $\chi^2(1, N = 900) = 5.26$, $p = .02$, $w=0.07$[62]. Inspection of the Pearson residuals (given in Table 20) indicates that the row 'abstract' is responsible for the minor effect – the number of abstract instances with *peal* is greater (1.49) and the number of abstract instances with adessive is smaller (-1.49) than expected by chance.

[62] As a reminder, the values of *w* for small, medium and large ES are .10, .30, and .50 respectively.

Table 20. Pearson residuals for Table 19

	adessive	peal
abstract	–1.49	1.49
spatial	0.64	–0.64

This result does not confirm the general assumption made in the literature about cases being more abstract than adpositions. My data show that when there is a meaning transfer, the adpositional construction is preferred. A similar result was obtained by Kesküla (2011: 35) who conducted a corpus study of the alternation between the Estonian allative and adposition *peale* 'onto' – the adpositional construction occurred significantly more frequently in abstract relations than the case construction. However, one must bear in mind here that only inter-changeable tokens of the adessive case were considered for the present analysis. If we compare the general usage of these two constructions, we can easily see that the adessive case expresses a large number of abstract functions where the use of an adposition is not possible (cf. sections 2.1.1 and 3.1.1). This, in turn, confirms the prediction made in literature that the case constructions are indeed more abstract than adpositional constructions. It might be the case then that when both constructions are applicable, language users prefer the adpositional construction for reasons of ambiguity. Using the adessive construction does not allow the speaker/writer to specify that they have in mind the metaphorical or metonymical locative function of the adessive and not one of the many 'truly' abstract functions like expressing the logical subject or time. It is interesting to note that if we consider the instances not as 'abstract' but 'figurative', then the results confirm those found in Mondorf (2003: 290) – she showed that an analytic comparative construction is preferred over synthetic comparative constructions with expressions of figurative meanings. Mondorf (*ibid*.) argues that this is indicative of a greater cognitive effort involved in denoting abstract or figurative concepts, which is compensated by the analytic construction, i.e. what she has termed *more*-support.

4.1.2 Type of Landmark

This variable has two levels – 'place' and 'thing' – and it is predicted that the adpositional construction is preferred with things as Landmarks, and the adessive construction with places as Landmarks. The frequency counts of this variable are given in Table 21. From Table 21 it can be seen that there was a more or less equal number of lexemes denoting places (449) as there were lexemes denoting things (451). The adessive construction is more frequent with places as Landmarks (265 out of 449) than the adpositional construction (184 out of 449). With things as Landmarks, the situation is the reverse – the adpositional construction is more frequent (266 out of 451) than the case construction (185 out of 451).

Table 21. Type of Landmark and construction (expected frequencies in parentheses)

Landmark	adessive	peal	Total
place	265 (224.5)	184 (224.5)	449
thing	185 (225.5)	266 (225.5)	451
Total	450	450	900

The results show that the frequency counts of the adessive and the adposition *peal* 'on' differ significantly by the type of Landmark, although the effect size is relatively small ($\chi^2(1, N = 900) = 29.16$, $p < .001$, w=0.18. Pearson residuals (presented in Table 22) show that both places and things contribute significantly for this difference – the number of places is significantly lower than expected with *peal* than with the adessive, and vice versa for things as Landmarks.

Table 22. Pearson residuals for Table 21

	adessive	peal
place	2.70	–2.70
thing	–2.69	2.69

As predicted, the adessive tends to be preferred when the Landmark is a place and the adposition *peal* 'on' when it is a thing. The same result was obtained by Kesküla (2011: 39) who studied the alternation between another pair, the allative and the adposition *peale* 'onto' – when the Landmark is a thing, the analytic *peale*-construction is preferred and when the Landmark is a place, the synthetic allative construction is preferred.

In order to see if there are any more subtle differences between the two constructions as pertains to the type of Landmark, the data in Table 21 is broken down into smaller groups according to 7 Landmark categories: abstract, body-part, building, clothing, furniture, place, and object. The results are presented in Table 23, from where it can be seen that the majority of Landmarks that occur in the adessive or *peal*-constructions are places (37%, 329 out of 900). The second biggest group is comprised of small objects (21%, 187 out of 900). These two groups are followed by furniture (12%, 112 out of 900), buildings (12%, 107 out of 900), and bodyparts (10%, 89 out of 900). In Table 21 above furniture, body parts, clothing, abstract Landmarks and small objects are collapsed, buildings and places are collapsed.

Table 23. Specific types of Landmarks and locative construction

Landmark	adessive	peal	Total	%
abstract	16	38	54	6%
bodypart	39	50	89	10%
building	67	40	107	12%
clothing	10	12	22	2%
furniture	50	62	112	12%
place	194	135	329	37%
small_object	74	113	187	21%
Total	450	450	900	100%

For clarity and a better overview of the data, information given in Table 23 is presented visually in Figure 1, where the types of Landmark are arranged not in alphabetical order as in Table 23, but according to the size of the Landmark – from relatively small Landmarks (small objects) to relatively large Landmarks (places). Abstract Landmarks do not fit very well onto this picture; they are presented in Figure 1 as the first category.

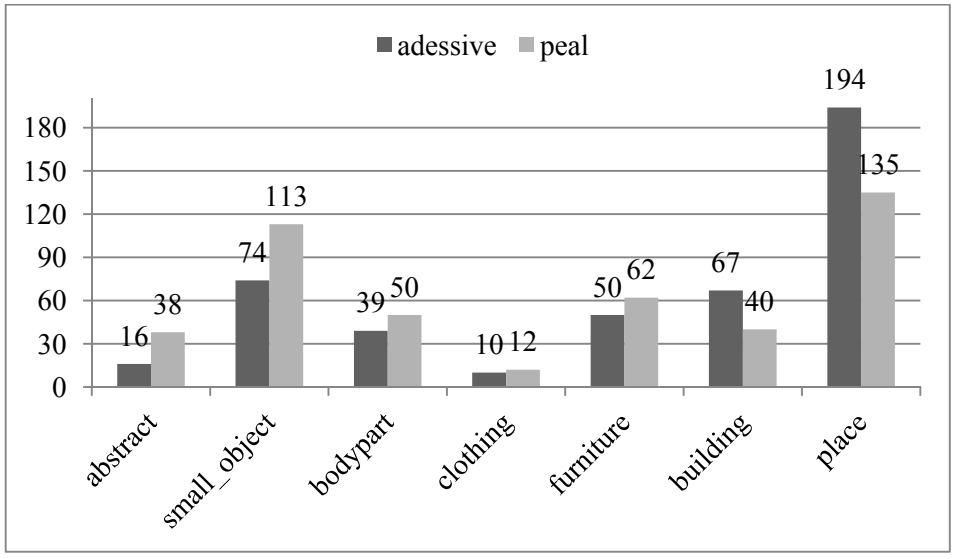


Figure 1. Frequency counts of specific type of Landmark with the two constructions ($N = 900$)

Table 21 above showed that there is a significant difference between the two constructions if we subcategorise the Landmarks into just two groups (i.e. things and places). Table 23 and Figure 1, however, indicate that there seem to be certain types of Landmarks where there is no difference between the two constructions. The groups that do not prefer one construction over the other

include clothes, bodyparts and furniture – the dark and light grey bars are of more or less equal height in Figure 1. Abstract Landmarks and small objects prefer the adpositional construction (the light grey bars are higher for these groups), while buildings and places prefer the case construction (dark grey bars are higher for these groups). This result is significant and has an even stronger effect size, although still small, compared to the results in Table 21 – $\chi^2(6, N = 900) = 37.32$, $p < .001$, $w = 0.20$. The Pearson residuals (given in Table 24) indicate that the groups of abstract Landmarks, buildings, places and things are responsible for this effect.

Table 24. Pearson residuals for Table 23

	adessive	peal
abstract	–2.12	2.12
bodypart	–0.82	0.82
building	1.85	–1.85
clothing	–0.30	0.30
furniture	–0.80	0.80
place	2.30	–2.30
small_object	–2.02	2.02

Table 25 presents the fifteen most frequent lexemes with the two constructions. The first column gives the specific lexeme and the second column indicates the exact frequency count. The Landmarks in bold (*laud* 'table', *trepp* 'stairs, *uks* 'door', and *lava* 'stage') indicate the lexemes that are included in both lists. These lexemes point to the fact that although in general there is a significant effect for the type of Landmark in the choice between the two constructions, i.e. that things prefer the *peal*-construction and places prefer the adessive, there are still some specific nouns that occurred equally frequently with both constructions. This is also mirrored in Table 23 and Figure 1. Another aspect to point out is that as seen from the frequency counts provided for *peal* 'on', there are two pronouns in the top five – *see* 'this' and *mis* 'what'. I will return to this point in the second part of this section, when I discuss the morphosyntactic variable 'word class of Landmark'. There were 192 different lexemes used with the *peal*-construction and 191 different lexemes used with the adessive construction. 73 different lexemes figured in both frequency lists. Such lexical effects are studied in detail in the second corpus study – section 4.4 presents the results of a distinctive collexeme analysis with the two Estonian locative constructions.

Table 25. 15 most frequent Landmarks with the adessive case and *peal* 'on' in multivariate corpus study

adessive	Count out of 450	*peal* 'on'	Count out of 450
tänav 'street'	17	**laud 'table'**	18
põrand 'floor'	13	see 'this'	15
laud 'table'	11	pilt 'picture'	14
rada 'path'	11	mis 'what'	13
trepp 'stairs'	11	hing 'soul'	11
ala 'area'	10	jalg 'foot'	9
aken 'window'	8	**trepp 'stairs'**	9
põld 'field'	8	õu 'yard'	7
katus 'roof'	7	põlv 'knee'	7
nägu 'face'	7	selg 'back'	7
piir 'border'	7	**uks 'door'**	7
saar 'island'	7	jää 'ice'	6
sein 'wall'	7	**lava 'stage'**	6
uks 'door'	7	paber 'paper'	6
lava 'stage'	6	tool 'chair'	6

4.1.3 Type of Trajector

There are two levels for the variable 'type of Trajector': 'abstract' and 'object'. No specific predictions are made as to effect of this variable on the choice between the adessive and the adpositional construction. No previous studies mention the Trajector, the focus has always been on the Landmark. The frequency counts for this variable are given in Table 26. It is clear from Table 26 that there is no difference between the two constructions as to the type of Trajector. Table 26 does, however, demonstrate that, overall, objects as Trajectors are much more frequent (846 out of 900) than abstract Trajectors (42 out of 900).

Table 26. Type of Trajector and construction (expected frequencies in parentheses)

Trajector	adessive	peal	Total
abstract	26 (27)	28 (27)	42
object	424 (423)	422 (423)	846
Total	450	450	900

The results of the chi-squared test confirm that there is no difference – $\chi^2(1, N = 900) = 0.08$, $p = 0.78$, $w = 0.01$.

4.1.4 Animacy of Landmark

This variable has two levels: 'animate' and 'inanimate'. It is predicted that when there is a need to express physical location of a Trajector on top of an animate Landmark, the preferred construction is the adpositional construction. Table 27 crosstabulates animacy with construction type. Overall, animate Landmarks are very infrequent – only 14 instances out of 900 in the dataset involved an animate Landmark. If the Landmark is animate, the preferred construction seems to be the adpositional construction (10 out of 14) and not the adessive (4 out of 14).

Table 27. Animacy of Landmark and construction (expected frequencies in parentheses)

Animacy of Lm	adessive	peal	Total
animate	4 (7)	10 (7)	14
inanimate	446 (443)	440 (443)	886
Total	450	450	900

However, the chi-squared test does not confirm this, the difference is not significant – $\chi^2(1, N = 900) = 2.61$, $p = .11$, $w = 0.05$. At the same time, Kesküla (2011: 37) reports in her study on the Estonian allative and adposition *peale* 'onto' alternation that animate Landmarks prefer the adpositional construction. These results demonstrate that very much as with the English genitive alternation, animacy is almost categorical in the alternation of Estonian locative constructions. 14 examples are clearly not enough to say anything conclusive about the effect of this variable in the present study and it requires further investigation.

4.1.5 Animacy of Trajector

Similarly to the previous variable 'animacy of Landmark', this variable has two levels: 'animate' and 'inanimate'. No specific predictions were made as to the preferred construction with either animate or inanimate Trajectors. The frequency counts are given in Table 28, where it can be seen that the distribution of animate and inanimate Trajectors is more or less equal between the two constructions. Overall, animate Trajectors are more frequent (525 out of 900) than inanimate Trajectors (375 out of 900).

Table 28. Animacy of Trajector and construction (expected frequencies in parentheses)

Animacy of Tr	adessive	peal	Total
animate	263 (262.5)	262 (262.5)	525
inanimate	187 (187.5)	188 (187.5)	375
Total	450	450	900

The results of a chi-squared test confirm that there is no significant effect for this variable – $\chi^2(1, N = 900) = 0.01$, $p = .95$, $w = 0.00$. It is clear from these results that similarly to the type of Trajector, the animacy of Trajector does not play a role in the choice between the adessive and *peal* either. A similar result was also found in the study on the Estonian allative case and the adposition *peale* 'onto' (Kesküla 2011: 37).

4.1.6 Mobility of Landmark

This variable has two levels: 'mobile' and 'static'. It is predicted that mobile Landmarks prefer the adpositional construction and static Landmarks the adessive construction. The frequency counts are given in Table 29. From this table it can be seen that, overall, there is a more or less equal number of mobile and static Landmarks – 438 and 462 out of a total of 900 respectively. Table 29 also indicates that there is a difference as to the mobility of Landmark and the type of locative construction – the adessive frequently occurs with a static Landmark and the adposition with a mobile Landmark.

Table 29. Mobility of Landmark and construction (expected frequencies in parentheses)

Mobility of Lm	adessive	peal	Total
mobile	170 (219)	268 (219)	438
static	280 (231)	182 (231)	462
Total	450	450	900

There is a moderate significant effect of this variable – $\chi^2(1, N = 900) = 42.71$, $p < .001$, $w = 0.22$. The Pearson residuals given in Table 30 indicate that the number of mobile Landmarks with the adessive construction is significantly lower (-3.31) than expected by chance, and that the number of static Landmarks with the adposition *peal* is also significantly lower (-3.22) than expected by chance. There is converging evidence for a similar effect of this variable in the alternation between the Estonian allative and *peale* 'onto' – static Landmarks prefer the allative case and mobile Landmarks the adposition *peale* 'onto' (Kesküla 2011: 42).

Table 30. Pearson residuals for Table 29

	adessive	peal
mobile	–3.31	3.31
static	3.22	–3.22

4.1.7 Mobility of Trajector

Similarly to the variable 'mobility of Landmark', this variable has also two levels: 'mobile' and 'static'. No specific predictions were made as to the preferred construction with either mobile or static Trajectors. The frequency counts are given in Table 31, where it can be seen that the distribution of animate and inanimate Trajectors is more or less equal between the two constructions. Overall, as is to be expected, mobile Trajectors are much more frequent (794 out of 900) than static Trajectors (106 out of 900).

Table 31. Mobility of Trajector and construction (expected frequencies in parentheses)

Mobility of Tr	adessive	peal	Total
mobile	398 (397)	396 (397)	794
static	52 (53)	54 (53)	106
Total	450	450	900

As with the results pertaining to other aspects of Trajectors, i.e. type of Trajector as well as animacy of Trajector, Table 31 confirms that the mobility of Trajectors does not have an effect on the choice between the two locative constructions – $\chi^2(1, N = 900) = 0.04$, $p = .84$, $w = 0.01$). However, we can conclude that static Trajectors are very infrequent with both the adessive and the *peal*-construction. This finding confirms the claim made by Talmy (2000: 315–316) that Figures or Trajectors are more movable than the Ground or Landmark.

4.1.8 Number of Landmark

This variable has two levels – 'plural' and 'singular' – and it is predicted that plural Landmarks prefer the adessive case construction. The overall frequency counts are given in Table 32. Overall, it can be seen that the number of plural Landmarks (110 out of 900) is low compared to singular Landmarks (790 out of 900). The number of plural Landmarks with the adessive case (67 out of 110) is larger than the number of plural Landmarks with the adposition *peal* (43 out of 110).

Table 32. Number of Landmark and construction (expected frequencies in parentheses)

Number of Lm	adessive	peal	Total
plural	67 (55)	43 (55)	110
singular	383 (395)	407 (395)	790
Total	450	450	900

Although the differences present in Table 32 are significant, there is only a very small significant effect: $\chi^2(1, N = 900) = 5.97$, $p = .01$, $w = 0.08$. The inspection

of the Pearson residuals given in Table 33 indicates that the proportion of plural Landmarks with *peal* is lower (-1.61) than expected and the proportion of plural Landmarks with adessive higher (1.61) than expected. According to the results presented in Kesküla (2011: 29–30), the number of the Landmark did not play a significant role in the choice between the allative case and *peale* 'onto'. Given the results of Kesküla (2011) and that the present study found only a very small effect for this variable, the number of Landmark might be considered as a potential variable to be included in further experimental studies.

Table 33. Pearson residuals for Table 32

	adessive	peal
plural	1.61	–1.61
singular	–0.60	0.60

4.1.9 Number of Trajector

Similarly to the number of Landmark, this variable has two levels: 'plural' and 'singular'. No specific predictions were made as to the preferred construction with either plural or singular Trajectors. The frequency counts are given in Table 34, where it can be seen that the distribution of animate and inanimate Trajectors is more or less equal between the two constructions. Overall, as with the number of Landmarks, singular Trajectors are more frequent (671 out of 900) than plural Trajectors (229 out of 900). Table 34 indicates that there seems to be a difference in the use of the adessive and *peal*-constructions according to whether the Trajector is singular or plural. The proportion of plural Trajectors with the adessive (132 out of 229) is higher than the proportion of plural Trajectors with *peal* (97 out of 229).

Table 34. Number of Trajector and construction (expected frequencies in parentheses)

Number of Tr	adessive	peal	Total
plural	132 (114.5)	97 (114.5)	229
singular	318 (335.5)	353 (335.5)	671
Total	450	450	900

The results of a chi-squared test confirm this, there is a significant, although a very small effect – $\chi^2(1, N = 900) = 7.18$, $p = .01$, $w = 0.09$. The inspection of the Pearson residuals given in Table 35 indicates that the proportion of plural Landmarks with *peal* is somewhat lower (–1.64) than expected and the proportion of plural Landmarks with adessive significantly higher (1.64) than expected. Kesküla (2011: 29) reports that there is a similar, very small

significant effect of this variable in the alternation between the allative and *peale* 'onto' – plural Trajectors prefer the case construction.

Table 35. Pearson residuals for Table 34

	adessive	peal
plural	1.64	–1.64
singular	–0.96	0.96

One of the reasons why plural Trajectors may prefer the synthetic case construction might be that there is an interaction with another variable – type of Landmark[63]. As was seen from Table 21 above which presented the results for the variable 'type of Landmark', places such as fields or streets as Landmarks prefer the adessive construction. One would expect to find more than one Trajector located in large places, e.g. many people or cars on a street vs. just one person or one car on the street. Table 36 crosstabulates the 132 instances of the adessive construction where the Trajector was plural and the type of Landmark. It can be seen that just over half of the adessive occurrences with plural Trajectors (76 out of 132, 58%) co-occur with 'places' as Landmarks.

Table 36. Adessive constructions with a plural Trajector and type of Landmark

Type of Landmark	
place	76 (58%)
thing	56 (42%)
Total	132 (100%)

4.1.10 Relative size of Trajector in relation to Landmark

This semantic variable has three levels: 'conventional', 'same', and 'unconventional'. 'Conventional' refers to situations where the Trajector is smaller than the Landmark; 'same' to situations where the two entities are of more or less the same size; and 'unconventional' to situations where the Trajector is bigger than the Landmark. The overall frequency counts are given in Table 37. The results indicate that in general, Landmarks tend to be larger than Trajectors, as predicted – the relative size is 'conventional' in 509 instances out of 900. This finding confirms the claims put forward by Talmy (2000: 315–316): Figures or Trajectors are smaller than the Ground or Landmark. Moreover, there is a difference between the adessive and the adposition *peal* 'on' – the adessive is preferred when the relative size of the Trajector in relation to Landmark is conventional (296 out of 509) and the adposition *peal* 'on' when the Trajector

[63] I am indebted to Krista Ojutkangas for this suggestion (p.c.).

and the Landmark are of the same size (139 out of 221) or when the Trajector is bigger than the Landmark (99 out of 170).

Table 37. Relative size of the Trajector in relation to the Landmark and construction (expected frequencies in parentheses)

Relative size of Tr and Lm	adessive	peal	Total
conventional	296 (254.5)	213 (254.5)	509
same	83 (110.5)	138 (110.5)	221
unconventional	71 (85)	99 (85)	170
Total	450	450	900

This difference is significant and has a moderate effect size: $\chi^2(2, N = 900) = 31.83$, $p < .001$, $w=0.19$. The Pearson residuals indicate that the rows 'conventional' and 'same' are responsible for the effect – adessive construction is used with 'conventional' size relations more often than expected by chance (2.60) and the adpositional construction with 'same' size relations (2.61). Kesküla (2011: 43) reports a similar finding – when the size between Trajector and Landmark is conventional the allative case is clearly preferred, and when the two entities are of the same size or when the Trajector is bigger than the Landmark, the adposition *peale* 'onto' is preferred.

Table 38. Pearson residuals for Table 37

	adessive	peal
conventional	2.60	–2.60
same	–2.61	2.61
unconventional	–1.52	1.52

4.1.11 Verb lemma

The first step in coding this variable involved determining the verb lemma used with each adessive and adpositional construction. In total, there were 277 different verbs in the dataset. The next step in the coding involved subcategorising the verb lemmas into 5 different groups: 'action', 'existence', 'motion', 'posture', and 'no verb'. The raw frequencies of these verb groups are given in Table 39. Overall, the most frequent verb group is that of action verb – 35% of verbs (318 out of 900) are action verbs. This group also includes the largest number of different verbs. It can be seen from Table 39 that the adessive case seems to be preferred with action verbs (179 out of 318). The next largest group of verbs includes existence verbs (22 % of verbs; 200 out of 900) like *olema*

'be', *asuma* 'be situated', *asetsema* 'be placed' and other verbs expressing existence or general location. Table 39 indicates that the adpositional construction seems to be preferred with existence verbs (126 out of 200). The third largest category – 15% of verbs (134 out of 900) concern verbs like *istuma* 'sit', *seisma* 'stand', *lebama* 'lie', *rippuma* 'hang' and other posture verbs. The fourth group of verbs comprises motion verbs (13%, 119 out of 900). In 129 instances out of 900, there was no verb expressed in clauses where the adessive and *peal*-constructions was used.

Table 39. Verb group and construction (expected frequencies in parentheses)

Verb group	adessive	peal	Total	%
action verbs	179 (159)	139 (159)	318	35%
existence verbs	74 (100)	126 (100)	200	22%
motion verbs	64 (59.5)	55 (59.5)	119	13%
posture verbs	64 (67)	70 (67)	134	15%
no verb	69 (64.5)	60 (64.5)	129	14%
Total	450	450	900	100%

The chi-squared test reveals that there is a small significant effect of verb – $\chi^2(4, N = 900) = 20.13$, $p < .001$, $w = 0.15$. The Pearson residuals given in Table 40 show that existence verbs determine the difference between the frequency counts – the adpositional construction occurred more often with existence verbs than expected by chance (2.60).

Table 40. Pearson residuals for Table 39

	adessive	peal
action verbs	1.59	–1.59
existence verbs	–2.60	2.60
motion verbs	0.58	–0.58
posture verbs	–0.37	0.37
no verb	0.56	–0.56

4.1.12 Summary of semantic variables

Table 41 presents the summary of the results of the univariate analysis of the semantic explanatory variables. The first column gives the variable name, the second and third column indicate which level of the variable is significantly associated with the specific locative construction, the fourth column gives the *p*-values, and the last column the effect size. The table is sorted according to the last column – from the largest effect size to the smallest. Table 41 is divided into three sections – the first section includes variables that are significant; the

second section includes variables that have a very small significant effect size and for which the Pearson residuals fall below +/-2; and the third section includes variables that are not significant. As mentioned above in section 3.3.3, the sample size (N = 900) of this corpus study is large enough to detect both medium and large effects. However, the largest effect size in this dataset – for the variable 'mobility of Landmark' – is only 0.22. Recall that Cohen (1992: 158, Table 2) reports .10 may be considered as a small effect, .30 medium, and .50 large. The conclusion to be drawn is that although 4 out 11 variables are significant, neither of these exhibit a particularly large effect. Overall, the properties of Landmarks (mobility, type, and number, but not animacy) seem to play at least some role, while as the properties of Trajectors (animacy, type, mobility) do not play any role at all in the choice between the two locative constructions. In general, the findings given here confirm the previous studies (cf. Bartens 1978, Ojutkangas 2008): case constructions are preferred with larger, static places and adpositional constructions with smaller, manipulable and mobile things. One of the puzzling findings pertains to the variable 'verb group' – existence verbs prefer the *peal*-construction. I return to this point at the end of this section, after I have presented the univariate results for morphosyntactic explanatory variables as well.

Table 41. Summary of the univariate results for semantic explanatory variables in multivariate corpus study

Variable name	Level for adessive	Level for peal	p-value	Effect size
mobility of LM	static	mobile	< .001	0.22
rel. size btw. TR&LM	TR > LM	TR = LM	< .001	0.19
type of LM	place	thing	< .001	0.18
verb group	-	existence	< .001	0.15
number of TR	(plural)	-	.01	0.09
number of LM	(plural)	-	.01	0.08
type of rel. btw. TR&LM	-	(abstract)	.02	0.07
animacy of LM	-	-	.11	0.05
type of TR	-	-	.78	0.01
mobility of TR	-	-	.84	0.01
animacy of TR	-	-	.95	0.00

4.1.13 Length of the Landmark phrase

For the first morphosyntactic variable, length, two different measures are used – length in words and length in syllables. Two conflicting predictions are made. Taking length as proxy for complexity, it can be predicted based on Mondorf (2003: 253) and Rohdenburg (2003: 205) that the analytical adpositional construction is preferred with longer Landmarks; alternatively, based on the

principle of language economy, the more compact synthetic adessive construction is preferred in longer environments. Figure 2 presents the results for the length of the Landmark phrase in words by the two locative constructions. Overall, the majority of the Landmark phrases are 1 (466 out of 900) or 2 (306 out of 900) words long. It is relatively rare to find a Landmark phrase that is 5 words long or longer in this dataset. The longest Landmark phrases are 7 words long and occur with the adessive case. From Figure 2 it can be seen that when the Landmark phrase is one word long, the more frequent construction is *peal*; for longer Landmark phrases, the adessive construction is more frequent.

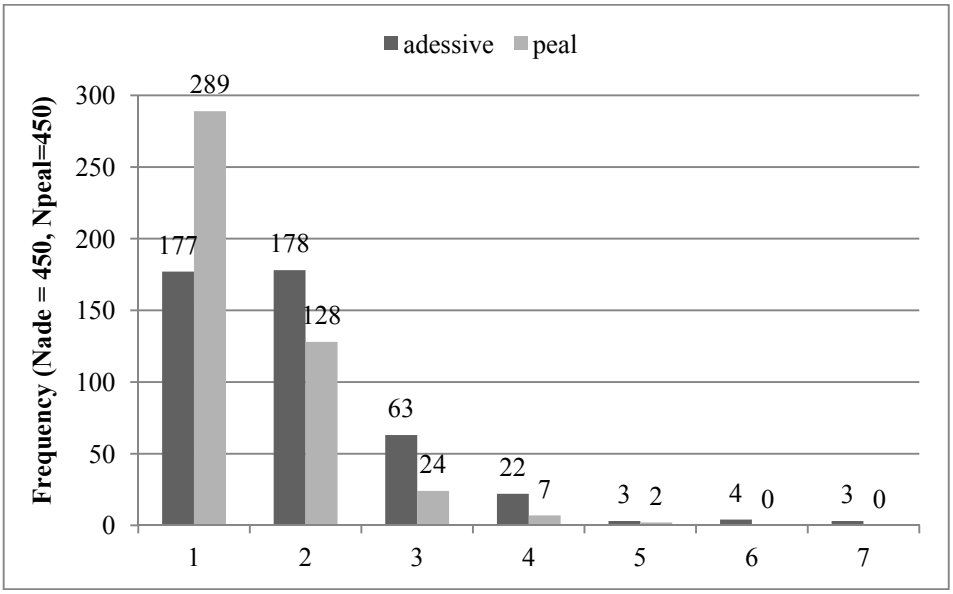


Figure 2. Frequency of Landmark length in words for adessive and *peal*-sentences

In order to interpret the results, the means, standard deviations, maximum and minimum values have been calculated based on the actual length; these are given in Table 42. The longest Landmark phrase with the adessive construction is 7 words long and with *peal* 5 words long. The shortest Landmark phrase in both constructions is 1 word long. The overall mean length of the Landmark phrase is 1.69 words (SD = 0.92). On average, the Landmark phrase in the adessive construction is longer than the Landmark phrase in the *peal*-construction. The mean length of the adessive Landmark phrases is 1.93 words (SD = 1.07) and *peal*-phrases 1.46 words (SD = 0.71). According to the *t*-tests for independent samples this difference is statistically significant with a medium effect size: $t = 8.02$, $df = 789.241$, $p < .001$, $d = 0.52$.

Table 42. Mean length of the Landmark phrase in words for the two constructions

Length in words	mean	SD	max.	min.
adessive	1.93	1.05	7	1
peal	1.46	0.71	5	1

Figure 3 presents the results of phonological complexity (length of the Landmark phrase in syllables) and type of locative construction. Overall, the majority of the Landmark phrases are 2 syllables long (305 out of 900). It is relatively rare to find a Landmark phrase that is 10 syllables long or longer in this dataset. The longest Landmark phrase is 41 syllables long and occurs with the adessive case. From Figure 3 it can be seen that when the Landmark phrase is one or two syllables long, the more frequent construction is *peal*; for longer Landmark phrases, the adessive construction is more frequent.

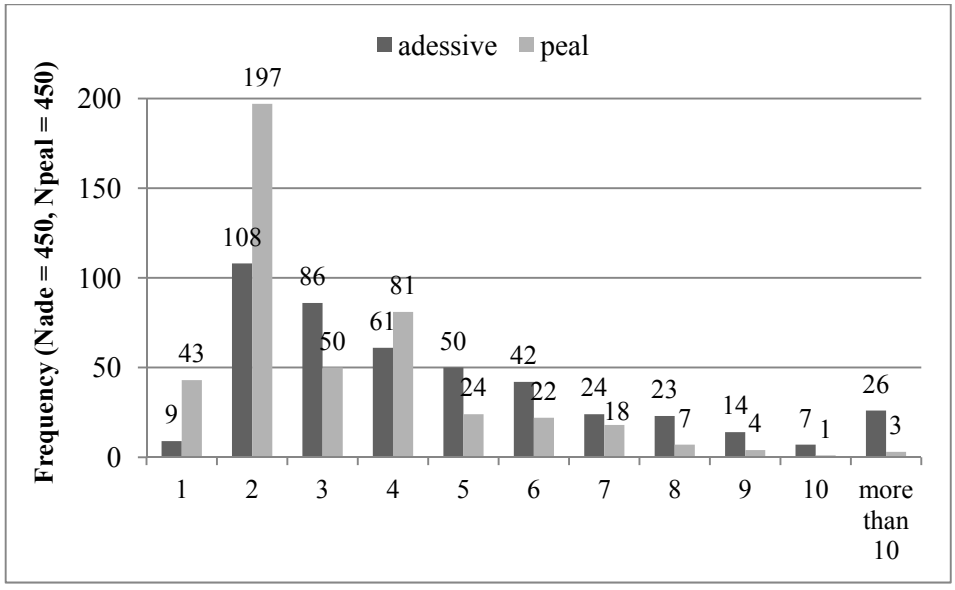


Figure 3. Frequency of Landmark length in syllables for adessive and *peal*-sentences

In order to interpret the results, the means, standard deviations, maximum and minimum values have been calculated based on the actual length; these are given in Table 43. The longest Landmark phrase with the adessive construction is 41 syllables long and with *peal* 11 syllables long. The shortest Landmark phrase in both constructions is 1 syllable long. The overall mean length of the Landmark phrase is 3.98 syllables (SD = 2.95). On average, the Landmark phrase in the adessive construction is longer than the Landmark phrase in the *peal*-construction. The mean length of the adessive Landmark phrases is 4.79 syllables (SD = 3.54) and *peal*-phrases 3.16 syllables (SD = 1.88). According to

the *t*-tests for independent samples this difference is statistically significant with a medium effect size: $t = 8.61$, $df = 682.177$, $p < .001$, $d = 0.58$.

Table 43. Mean length of the Landmark phrase in syllables for the two constructions

Length in words	mean	SD	max.	min.
adessive	4.79	3.54	41	1
peal	3.16	1.88	11	1

As we saw in the theoretical discussion in section 2.3.2, both length measures of the Landmark phrase – number of words and number of syllables – can be said to reflect complexity. The first measure is associated with syntactic complexity and the second with phonological complexity. If we take length as proxy of complexity and proceed from Mondorf's (2003: 253) analytic support, then the prediction that the analytic *peal*-construction will be used with longer (i.e. more complex) Landmark phrases does not hold for the Estonian locative alternation according to the two corpus studies. It is instead the synthetic adessive construction that is preferred with longer Landmark phrases. Hence, it is plausible to presume that instead of the analytic support, the principle of economy plays a role – language users prefer the more economical synthetic construction with longer phrases. Another explanation why the adessive construction is preferred with longer Landmarks might have to do with explicitness. Length of the Landmark phrase may be taken as a proxy for explicitness or specificity and it may be concluded that the more specific adpositional construction is used to compensate for the relatively low degree of explicitness of the Landmark phrase itself; when the Landmark itself is explicit enough, i.e. long, the less explicit locative construction, i.e. the locative case, is used. The results of the sentence rating task described in Chapter 5 examine the issue of length of the Landmark phrase and the choice between the two constructions further.

4.1.14 Morphological complexity of Landmark

This variable has two levels: 'compound' and 'simple'. Morphological complexity of the Landmark is taken as proxy for complexity and it is predicted, based on Mondorf's (2003: 205) analytic support, that the analytic adpositional construction is used with more complex Landmarks. The overall frequency counts are given in Table 44, from where it can be seen that in general, the majority of the Landmarks are simple lexemes – 733 out of 900. Table 44 also shows that the proportion of adessive constructions with morphologically complex Landmarks is considerably larger than the proportion of *peal*-constructions – as many as 133 out of 167 compounds are used in the adessive case.

Table 44. Morphological complexity of the Landmark and construction (expected frequencies in parentheses)

Number of Tr	adessive	peal	Total
compound	133 (83.5)	34 (83.5)	167
simple lexeme	317 (366.5)	416 (366.5)	733
Total	450	450	900

There is a strong significant effect of this variable: $\chi^2(1, N = 900) = 72.06$, $p < .001$, $w = 0.28$. The Pearson residuals in Table 45 indicate that the adessive is used with compound lexemes considerably more often than expected by chance (5.42). This gives further indication of the adposition *peal* being preferred with shorter, less complex and less explicit Landmark phrases, while the adessive tends to be used with longer, more complex and more explicit Landmark phrases.

Table 45. Pearson residuals for Table 44

	adessive	peal
compound	5.42	–5.42
simple lexeme	–2.59	2.59

4.1.15 Syntactic function of the Landmark phrase

This variable has two levels: 'adverbial' and 'modifier'. It is predicted that the adessive case is the preferred construction when the Landmark phrase is used in the modifier function. Table 46 gives the overall frequencies – both the adessive and *peal*-construction predominantly occur in the adverbial function; there are only 87 instances (out of 900) where the locative construction functions as a modifier. There seems to be a difference between the two constructions in the syntactic functions they fulfil in a clause – the adessive construction (57 out of 87) occurs in the modifier function more often than *peal*-construction (30 out of 87).

Table 46. Syntactic function of the Landmark phrase and construction (expected frequencies in parentheses)

Syn. function	adessive	peal	Total
adverbial	393 (406.5)	429 (406.5)	813
modifier	57 (43.5)	30 (43.5)	87
Total	450	450	900

Although the differences highlighted in Table 46 are significant [$\chi^2(1, N = 900) = 9.28$, $p = .002$, $w = 0.10$], the effect size of this variable is small. The Pearson residuals given in Table 47 indicate that the adessive is used in the modifier function more frequently than expected by chance (2.05).

Table 47. Pearson residuals for Table 46

	adessive	peal
adverbial	–0.66	0.66
modifier	2.05	–2.05

4.1.16 Word class of Landmark

Word class of Landmark has two levels – 'noun' and 'pronoun' – and it is predicted that the adpositional construction is preferred with pronominal Landmarks. The frequency counts presented in Table 48 show that the majority of the Landmark phrases inflected for the adessive case or occurring in the adpositional phrase with *peal* are full noun phrases (869 out of 900); only 40 Landmarks are pronominal.

Table 48. Word class of Landmark and construction (expected frequencies in parentheses)

Word class of Lm	adessive	peal	Total
noun	442 (430)	418 (430)	860
pronoun	8 (20)	32 (20)	40
Total	450	450	900

The prediction that there is an association between word class of the Landmark and construction type is confirmed – $\chi^2(1, N = 900) = 15.07$, $p < .001$, $w = 0.13$. Although significant, the effect size for this variable is small. The Pearson residuals in Table 49 indicate that the prediction that when the Landmark is a pronoun, the *peal*-construction is preferred is confirmed. The results for this variable are related to the other morphosyntactic variable 'length of the Landmark phrase' – pronouns, in addition to being less specific than full nouns, are also shorter than full nouns and these results reflect the tendency to use the adpositional construction with shorter Landmarks and the case construction with longer Landmarks. Kesküla (2011: 31) provides converging evidence as well – there is a very strong preference for the adposition *peale* 'onto' to be used with pronouns.

Table 49. Pearson residuals for Table 48

	adessive	peal
noun	0.57	–0.57
pronoun	–2.68	2.68

4.1.17 Word class of Trajector

Trajectors may belong to one of the following three word classes – 'noun', 'pronoun', and 'verb phrase'. No specific predictions were made as to the preferred construction with either plural or singular Trajectors. The frequency counts are given in Table 50, where it can be seen that in general, as with the word class of Landmarks, nominal Trajectors are more frequent (586 out of 900) than pronominal Trajectors (166 out of 900) or verbal Trajectors (148 out of 900). Table 50 indicates that there seems to be a difference in the construction type according to the word class of Trajector. The proportion of nominal Trajectors with the adessive (324 out of 586) is higher than the proportion of nominal Trajectors with *peal* (262 out 586), while the proportion of pronominal Trajectors with *peal* (104 out of 166) is higher than the proportion of pronominal Trajectors with the adessive (62 out 166). With verbal Trajectors, both constructions are used with more or less equal frequency, with *peal* occurring slightly more frequently (84 out of 148).

Table 50. Word class of Trajector and construction (expected frequencies in parentheses)

Word class of Tr	adessive	peal	Total
noun	324 (293)	262 (293)	586
pronoun	62 (83)	104 (83)	166
verb phrase	64 (74)	84 (74)	148
Total	450	450	900

There is a moderate significant effect for this variable: $\chi^2(2, N = 900) = 19.89$, $p < .001$, $w = 0.15$. The Pearson residuals in Table 51 indicate that it is the pronominality of the Trajector that determines the difference between the frequency counts – the adposition *peal* occurs with pronominal Trajectors (2.31) more often than expected by chance.

Table 51. Pearson residuals for Table 50

	adessive	peal
noun	1.81	–1.81
pronoun	–2.31	2.31
verb phrase	–1.16	1.16

4.1.18 Case form of Trajector

This variable has four levels: 'nominative', 'partitive', 'other cases', 'not applicable'. No specific predictions were made as to the preferred construction with Trajectors inflected for different cases. Table 52 presents the overall frequency counts and it can be seen that the nominative is the preferred case for the Trajector – a large majority of the Trajectors are in the nominative (583 out of 900). Still, a comparatively large proportion of Trajectors (124 out of 900) are also in the partitive case. It is relatively rare for the Trajector to be in any of the other case (44 out of 900) besides the nominative or the partitive.

Table 52. Case form of Trajector and construction (expected frequencies in parentheses)

Case form of Tr	adessive	peal	Total
nominative	287 (291.5)	296 (291.5)	583
partitive	71 (74.5)	53 (74.5)	124
other cases	24 (22)	20 (22)	44
not applicable	68 (62)	81 (62)	149
Total	450	450	900

Kesküla (2011: 28) shows that the case form of the Trajector is a significant variable in the alternation between the allative and *peale*-construction – when the Trajector is in the nominative, the case construction is preferred, and when the Trajector is in the partitive, the adpositional construction is preferred. The frequency counts in Table 52 present conflicting evidence – the number of adessive constructions with nominative Trajector (287 out of 583) is lower than the number of *peal*-constructions (296 out of 583). My data do not confirm Kesküla's (2011: 28) other finding either – when the Trajector is in the partitive the preferred construction is not the adpositional construction, but the case construction. However, these minor differences in the frequency counts in Table 52 are not significant: $\chi^2(3, N = 900) = 4.25$, $p = 0.24$, $w = 0.07$. At the same time, this variable is significant in Kesküla's (2011) study. Hence, no definite conclusions can be drawn about the role of this variable in the choice between locative cases and adpositions. It may very well be that this variable does play a role in the choice between the allative and *peale*, but not in the choice between the adessive and *peal*. Still, as Kesküla (2011: 28) points out, it may also be the case that this variable is not a significant variable of its own, but that the case form of the Trajector may indicate that some other aspect is determining the choice between the locative cases and adpositions which in turn then also affects the case form of the Trajector. It is hypothesised that one of these 'more important' variables might be the clause pattern; I discuss the results of this variable below.

4.1.19 Clause type

Each clause in which the adessive and *peal*-construction is used is coded for the variable 'clause type' which has two levels: 'main' or 'subordinate'. No specific predictions were made as to the preferred construction and the clause type. The overall frequency counts are given in Table 53, from where it can be seen that the majority of the clauses are main clauses (573 out of 900). The adessive case construction (308 out of 573) is more frequent with main clauses than the *peal*-construction (265 out of 573); the situation is reverse for subordinate clauses.

Table 53. Clause type and construction (expected frequencies in parentheses)

Type	adessive	peal	Total
main	308 (286.5)	265 (286.5)	573
subordinate	142 (163.5)	185 (163.5)	327
Total	450	450	900

This result is significant [$\chi^2(1, N = 900) = 8.88$, $p = .002$, $w = 0.10$], but the variable as a whole has a very small effect size. The Pearson residuals in Table 54 indicate that the *peal*-construction is only marginally more frequent with subordinate clauses than expected by chance. This result is interrelated with another morphosyntactic variable – that of 'word class of Landmark'. The results indicate that the adpositional construction is preferred with pronominal Landmarks, the majority of which occur in subordinate clauses.

Table 54. Pearson residuals for Table 53

	adessive	peal
main	1.27	–1.27
subordinate	–1.68	1.68

4.1.20 Position of the Landmark phrase

This word order variable has three levels: 'initial', 'middle', and 'final', depending on the position of the Landmark phrase in the clause. Based on the principle of end-weight which states that "long, complex phrases tend to come at the ends of clauses" (Wasow 1997: 81), I assume that it is the analytic adpositional construction with *peal* 'on' that can create a heavier constituent because it has the extra lexeme (*peal* 'on') and that the adpositional construction should thus prefer the clause-final position. This prediction finds further support in information structure – given that the adpositional construction is more explicit and specific than the case construction, it should be used at the end of clauses where new information is provided in Estonian.

The case construction is predicted to be used at the beginning of a clause because it is shorter than the adpositional construction and less specific. Table 55 presents the overall frequency counts. The most frequent position for both types of constructions is at the middle of the clause (398 out of 900), followed by the clause-final position (343 out of 900). It is relatively rare for the adessive or *peal*-construction to appear at the beginning of the sentence (249 out of 900).

Table 55. Position of the locative phrase within the clause and construction (expected frequencies in parentheses)

Position	adessive	peal	Total
final	182 (171.5)	161 (171.5)	343
initial	127 (124.5)	122 (124.5)	249
middle	141 (154)	167 (154)	398
Total	450	450	900

The distribution of the two locative constructions between the three positions seems more or less equal. This is confirmed by a chi-squared test, which shows that there is no significant effect for this variable: $\chi^2(2, N = 900) = 3.58$, $p = .17$, $w = 0.06$. One of the reasons why the prediction concerning the preference of the *peal*-construction to occur at the end of a clause is not confirmed might have to do with the variable 'length of the Landmark phrase'. As we saw above, the adessive case is preferred with longer Landmark phrases and longer Landmark phrases prefer the clause-final position.

4.1.21 Relative position of the Trajector and Landmark phrase

The second word order variable has two levels: 'lm_tr' and 'tr_lm'. This variable looks at the relative position of the Trajector and Landmark phrases. The code 'lm_tr' indicates that the Landmark phrase precedes the Trajector phrase, while 'tr_lm' indicates that it follows the Trajector phrase. The overall frequency counts, given in Table 56, confirm that in general the preferred word order is such that the Landmark phrase follows the Trajector phrase ('tr_lm'); 576 clauses out of 900 have this word order sequence.

Table 56. Relative position of the Trajector and Landmark phrase and construction (expected frequencies in parentheses)

Position	adessive	peal	Total
lm_tr	174 (162)	150 (162)	324
tr_lm	276 (288)	300 (288)	576
Total	450	450	900

Although it seems that the adessive construction rather than the *peal*-construction is preferred with the 'lm_tr' word order (174 vs. 150 out of 324), this difference is not significant: $\chi^2(1, N = 900) = 2.78$, $p = .09$, $w = 0.06$. The two rating tasks described in Chapter 5 address this issue in more detail.

4.1.22 Summary of morphosyntactic variables

Table 57 presents the summary of the results of the univariate analysis of the morphosyntactic explanatory variables. The table is divided into three sections – the first section includes variables that are significant; the second section includes variables that have a very small significant effect size and for which the Pearson residuals fall below +/-2; and the third section includes variables that are not significant. The first column in Table 57 specifies the variable name, the second and third column indicate which level of the variable is significantly associated with the specific locative construction, the fourth column gives the *p*-values, and the last column the effect size. The table is sorted according to the last column – from the largest effect size to the smallest.

Table 57. Summary of the univariate results for morphosyntactic explanatory variables in multivariate corpus study

Variable name	Level for adessive	Level for peal	p-value	Effect size
length of LM in syllables	mean=4.79 SD=3.54	mean=3.16 SD=1.88	< .001	0.58
length of LM in words	mean=1.93 SD=1.05	mean=1.46 SD=0.71	< .001	0.52
morphol. complex. of LM	compound	simple	< .001	0.28
word class of TR	–	pronoun	< .001	0.15
word class of LM	–	pronoun	< .001	0.13
clause type	–	(subord.)	.002	0.10
syntactic function of LM	(modifier)	–	.002	0.10
rel. position of TR&LM	–	–	.09	0.06
position of LM phrase	–	–	.17	0.06
case form of TR	–	–	.24	0.07

As mentioned above in section 3.3.3, the sample size ($N = 900$) of this corpus study is large enough to detect both medium and large effect sizes. As a reminder, .10 may be considered a small effect, .30 medium, and .50 large for chi-squared test; for the *t*-test, small, medium, and large effect sizes are .20, .50, .80 respectively (Cohen 1992: 158, Table 2). The largest effect size (0.58) in Table 57 is for the variable 'length of the Landmark phrase in syllables'; since the statistical test used for assessing the significance of this variable is the independent samples *t*-test, the effect may be considered as medium. The other

two medium effect sizes are detected for the variable 'length of the Landmark phrase in words' (0.52) and 'morphological complexity of the Landmark' (0.28). In essence, all three variables relate to one single phenomenon – the tendency for the synthetic adessive case to be preferred with longer and more complex phrases and the analytic adpositional construction with shorter and less complex phrases. This result indicates that contrary to the principle of end-weight (Hawkins 1994, 2004; Wasow 1997, 2002) and the analytic support proposed by Mondorf (2003), the principle of language economy plays a role here. In order to avoid making the already long and complex phrase even longer by adding an extra word *peal*, language users opt for the more concise adessive case inflection. The other two variables that are significant – the word class of both Trajector and Landmark – exhibit a small effect and point towards the conclusion that the adpositional construction is preferred with pronominal Landmarks and Trajectors and with subordinate clauses. This result is again related with length – pronouns tend to be shorter than full noun phrases.

4.2 Summary of the univariate results of Multivariate Corpus Study

Table 58 presents in a summary fashion all the explanatory variables studied in the first corpus study, the *p*-values and effect sizes obtained using the chi-squared test or a *t*-test for independent samples. The variables are ordered according to the effect size from the most significant variable to the least significant variable. The column titled 'Variable group' indicates whether a specific variable is a semantic or morphosyntactic variable. The columns titled 'Level for adessive' and 'Level for *peal*' indicate which level of the specific variable is responsible for the significant effect found in the frequency tables and with which construction the observed frequency is higher than expected by chance. The upper part of the table presents the significant results and the bottom part of the table lists the variables that were not significant. Similarly to Tables 41 and 57 above, Table 58 is divided into three sections – the first section includes variables that are significant; the second section includes variables that have a very small significant effect size and for which the Pearson residuals fall below +/-2; and the third section includes variables that are not significant.

Overall, nine variables out of a total of 20 variables are significant. Five of these significant variables are morphosyntactic and four are semantic. Out of the five variables that had a very small significant effect, three were semantic and two morphosyntactic. Out of the seven variables that were not significant, four were semantic and three morphosyntactic. Thus it seems that both semantic as well as morphosyntactic aspects are relevant in the alternation between Estonian locative cases and adpositions.

Table 58. Univariate results of multivariate corpus study

Variable name	Variable group	Level for adessive	Level for peal	p-value	Effect size
length of LM in syllables	morphosyn.	mean=4.79 SD=3.54	mean=3.16 SD=1.88	< .001	0.58 (medium)
length of LM in words	morphosyn.	mean=1.93 SD=1.05	mean=1.46 SD=0.71	< .001	0.52 (medium)
morphol. complex. of LM	morphosyn.	compound	simple	< .001	0.28 (medium)
mobility of LM	semantic	static	mobile	< .001	0.22 (medium)
rel. size btw. TR&LM	semantic	TR > LM	TR = LM	< .001	0.19 (small)
type of LM	semantic	place	thing	< .001	0.18 (small)
verb group	semantic	–	existence	< .001	0.15 (small)
word class of TR	morphosyn.	–	pronoun	< .001	0.15 (small)
word class of LM	morphosyn.	–	pronoun	.001	0.13 (small)
clause type	morphosyn.	–	(subord.)	.002	0.10 (small)
syntactic function of LM	morphosyn.	(modifier)	–	.002	0.10 (small)
number of TR	semantic	(plural)	–	.01	0.09 (small)
number of LM	semantic	(plural)	–	.01	0.08 (small)
type of rel. btw. TR&LM	semantic	–	(abstract)	.02	0.07 (small)
animacy of LM	semantic	–	–	.02	0.07
case form of TR	morphosyn.	–	–	.24	0.07
position of LM phrase	morphosyn.	–	–	.17	0.06
rel. position of TR&LM	morphosyn.	–	–	.09	0.06
type of TR	semantic	–	–	.78	0.01
mobility of TR	semantic	–	–	.84	0.01
animacy of TR	semantic	–	–	.95	0.00

There are only three variables with a relatively strong effect size – length of the Landmark phrase in syllables (d = 0.58) and words (d = 0.52), and morphological complexity (w = 0.28) in Table 58; the other significant variables exhibit a moderate effect size. The dataset is large enough to detect both medium and large effects, so it cannot be that there is not enough data. The fact that there are not very many variables that have a large effect size is not surprising given the complex and multivariate nature of the phenomenon. No single variable single-handedly determines the choice between the adessive and *peal*-constructions in Estonian. It may be the case that the individual significant variables identified in the dataset may influence the choice between the two locative constructions in combination. The next section looks specifically at this issue – it presents the multivariate analysis of the first corpus study.

The majority of the significant morphosyntactic variables have to do with complexity and length of the Landmark phrase. It is predicted, based on Mondorf's (2003) analytic support and the related issues, that the adpositional construction as the more explicit construction of the two is preferred with longer and more complex Landmark phrases. The results show a reverse result – it is the more concise adessive case that is preferred in cognitively more complex environments. It is plausible to claim that the principle of language economy plays a role instead – language users prefer the shorter synthetic

construction over the longer analytic construction with a Landmark phrase that is itself already long and complex.

Length issues may be the reason why the semantic variable 'verb group' has a moderately significant effect as well. According to the frequency counts in multivariate corpus study, the adpositional construction occurs frequently with existence verbs. One such frequent verb of existence is the copula *olema* 'be', which in the 3rd person singular is a relatively short word both in the present (*on* 'is') as well as in the past tense (*oli* 'was') and prefers the *peal*-construction. Another reason for this result may lie in the semantics of both the verb and the construction. On the one hand, verbs of existence are relatively empty and do not say much about the properties of the Landmark and Trajector and the type of relation between them. On the other hand, the adpositional construction has been claimed to be more specific or explicit than the case construction (Comrie 1986, Luraghi 1991: 66–67, Hagège 2010: 37–38, and Lestrade 2010). Put together, these two elements balance each other out – the adpositional construction provides enough (semantic) information and allows for a relatively empty verb to be used. With the case construction, the situation is reversed – because the adessive itself is relatively empty and fulfils many other functions besides expressing location, it requires the verbal component to be more specific and carry more meaning. By employing the *olema* + *peal*-construction, the language users focus more on the location of the trajector, while the use of the adessive case with other types of verbs the focus is on the activity (i.e. what the Trajector is doing and less how or where it is located). Hence, the different division of labour between the verb and the locative construction. A third possible explanation is that the language users wish to avoid ambiguity. The copula *olema* 'be' is also the verb used in the Estonian possessive construction with the possessor in the adessive case. Again, it is plausible that all of these explanations are applicable and influence the speakers' choice.

The three most significant semantic variables are mobility of the Landmark ($w = 0.22$), relative size of the Trajector in relation to the Landmark ($w = 0.19$), and type of Landmark ($w = 0.18$). In general, these results confirm the results found in previous studies in other Finno-Ugric languages (Bartens 1978, Ojutkangas 2008): the case construction is preferred with larger, static places and the adpositional construction with smaller, manipulable and mobile things. Even though the semantic variable 'number of Trajector' is only marginally significant ($w = 0.09$), the interpretation of why it should be the adessive construction that occurs with plural Trajectors more frequently than expected by chance may have to do with the above semantic variables as well. The adessive construction is used with larger Landmarks and larger Landmarks more readily allow plural Trajectors to be placed upon them. Hence, this result is another way of saying that the type of Landmark plays a significant role in the alternation between the adessive and adposition *peal* 'on'.

The two morphosyntactic variables of word order ('position of the Landmark phrase' and 'the relative position of Trajector and Landmark') did not have a significant effect in the choice between adessive and adposition *peal* – with

both constructions the preferred word order is such that the Trajector phrase precedes the Landmark phrase and both constructions prefer the clause-middle position. Although mobility and type of the Landmark phrase play a role, animacy and number do not. It may be the case that the number of animate Landmarks in the present dataset is not enough to detect any difference between the two constructions – humans or animals are not prototypical Landmarks on top of which Trajectors are placed. In comparison to the many properties of the Landmark, properties of the Trajector do not play a role in the alternation between the two constructions. The majority of the Trajectors are in the nominative case, mobile and manipulable things irrelevant of the type of locative construction.

In general, the univariate results of the first corpus study point towards the conclusion that the short synthetic case construction is used in morphosyntactically more complex environments, and the adpositional construction in less complex contexts. There is evidence that the case construction has a more abstract and less specific constructional meaning and the adpositional construction has a more specific constructional meaning. The univariate results of the two corpus studies permit the following observations about the usage patterns of the two locative constructions:

i) The **adessive** construction tends to be used with:
   - complex and long noun phrases as Landmarks;
   - static places as Landmarks;
   - Landmarks that are bigger than Trajectors.

ii) The ***peal***-construction tends to be used with:
   - short and simple noun phrases;
   - small, mobile things as Landmarks;
   - Landmarks that are of the same size as Trajectors;
   - verbs of existence;
   - pronominal Trajectors and Landmark.

4.3 Multivariate results of Multivariate Corpus Study

This section makes use of binary logistic regression analysis as a multivariate statistical technique for explaining the relationship between the binary response variable – type of locative construction – and the explanatory variables that were identified as having an effect on the response variable. Although the results of the univariate analysis indicate that 14 variables are significant in the alternation between the Estonian adessive and adposition *peal* 'on', this way of looking at the data is only one aspect of the bigger picture. When language users make use of either of these locative constructions, they probably do not consider the value of one variable alone – in actual language use, all of the variables interact simultaneously and need to be analysed as such. Therefore, a

multivariate approach is necessary to determine which of the variables are more decisive and predictive for the choice between the two constructions. However, given that all of the variables that are identified as significant by the univariate analysis have either a small or medium effect size (no large effect sizes are detected), it is to be assumed that it will not be possible to fit a particularly good and strong model to the data.

The univariate results (specifically the effect sizes and Pearson residuals) reported in the previous section are taken as the basis for model building. First, a model with only one explanatory variable that had the strongest effect size ('length of the Landmark phrase in syllables') was built. Then, a model with two explanatory variables that had the strongest effect sizes ('length of Landmark phrase in syllables' and 'mobility of Landmark')[64] was built, and so on. The fit of competing models is measured by comparing the respective deviance values and other model diagnostics (see section 3.3.1 above). An added variable is left in the model only if it is significant and improves the model fit. The following variables did not prove to be significant predictors in the multivariate analysis: relative size of the Trajector in relation to the Landmark, Landmark type, word class of Landmark, clause type, syntactic function of the Landmark phrase, number of Landmark and Trajector, relation type between Trajector and Landmark. None of these explanatory variables are significant on their own, nor do they participate in any significant interactions – including them in the model does not improve model fit. It can be concluded that although the univariate results show a moderate or small significant effect for some of these variables, they do not participate in a significant way in combination with the other variables in predicting the use of the two Estonian locative constructions.

It may be the case that the variables which are not reported as significant according to the univariate results, turn out to be significant once they are treated in combination with the other variables. To check this, different models were built by adding the seven insignificant variables. Five out of the seven variables are confirmed not to be significant for the multivariate analysis either: animacy of the Landmark and Trajector, case form of Trajector, type and mobility of Trajector. It turns out, however, that the two word order variables, 'position of the Landmark phrase' and 'the relative position of Trajector and Landmark', significantly improve the model fit when modelled in together with the variables 'length' and 'morphological complexity of the Landmark phrase', 'mobility of Landmark', 'verb group', and 'word class of Trajector'. It is important to keep in mind that it is wrong to claim that the variables that are dropped from the final models are not correlated with the response variable –

[64] The variable 'length of the Landmark phrase in words' exhibits the second strongest effect size, but the two length measures can not be used simulatenously in one single model due to reasons of multicollinearity. The variable 'length in syllables' is a stronger predictor than the variable 'length in words', i.e. the inclusion of the former liscenses better models than the inclusion of the latter. It is concluded that for the present dataset, length measured in syllables is a more adequate operationalisation of the variable length than length measured in words.

these variables just do not add enough explanatory effect over the other variables that are already in the model for it to become significantly better (Faraway 2002: 126).

Table 59 lists, in alphabetical order, the variables that are included in the minimally adequate model fitted to the multivariate corpus data. The first column gives the variable name. The second column gives the specific levels of each variable. The third column provides the code used for these levels in the logistic regression analysis and for each variable (except for the only numerical variable 'length'), the reference level used in the regression is also indicated.

Table 59. List of variables in binary logistic regression model (multivariate corpus study)

Variable name	Levels	Code
mobility of LM	mobile	LM_MOBILITYmobile = ref. level
	static	LM_MOBILITYstatic
verb group	action	VERB_GROUPaction = ref. level
	existence	VERB_GROUPexistence
	motion	VERB_GROUPmotion
	posture	VERB_GROUPposture
	no verb	VERB_GROUPnoverb
length of LM phrase (log. transformed)	in syllables	LM_LENGTHSYLLOG
morphol. complexity of LM	compound	LM_COMPcompound = ref. level
	simple	LM_COMPsimple
word class of TR	noun	TR_WCnoun = ref. level
	pronoun	TR_WCpronoun
	verb phrase	TR_WCVP
relative position btw TR&LM	lm_tr	WO_LMlm_tr = ref. level
	tr_lm	WO_LMtr_lm

The output for the minimally adequate model fitted to the multivariate corpus data is presented in Figure 4. It has the 'type of construction' (CONSTRUCTION) as the response variable and the following six as explanatory variables: 'length of the Landmark phrase in syllables', 'morphological complexity of the Landmark', 'mobility of the Landmark', 'verb group', 'word class of Trajector', and 'relative position of Trajector and Landmark phrase'. The second word order variable 'position of the Landmark phrase' is also a significant variable, but including it in the model together with the variable 'relative position of Trajector and Landmark' creates problems of multicollinearity. The variance inflation factors (VIF) for the level 'initial' of the variable 'position of the Landmark phrase' is 4.18 and for the level 'tr_lm' of the variable 'relative position of Trajector and Landmark' is 3.52. This indicates that these two variables are correlated and they are both trying to do the same job in explaining the response variable. It is therefore correct to not include these two word order variables simultaneously in the model. Building the model with the

variable ‘relative position of Trajector and Landmark’ leads to a stronger model than building the model with the variable ‘position of the Landmark phrase’. The model presented in Figure 4 therefore includes the variable ‘relative position of Trajector and Landmark’ and not the other word order variable. This, however, should not be taken to mean that the variable ‘position of the Landmark phrase’ is not significant – it is significant, but weaker than the variable ‘relative position between Trajector and Landmark’.

```
Call:
glm(formula = CONSTRUCTION ~ LM_LENGTHSYLLOG + LM_COMP +
LM_MOBILITY + VERB_GROUP + TR_WC + WO_LM, family = "binomial",
data = basic)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.14872  -0.96006   0.03036   0.94679   2.30562

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)           0.01344    0.34283   0.039 0.968731
LM_LENGTHSYLLOG      -0.92625    0.13590  -6.816 9.39e-12 ***
LM_COMPsimple         1.16446    0.22373   5.205 1.94e-07 ***
LM_MOBILITYstatic    -0.89143    0.15347  -5.808 6.30e-09 ***
VERB_GROUPexistence   0.70680    0.20904   3.381 0.000722 ***
VERB_GROUPmotion      0.03419    0.24192   0.141 0.887624
VERB_GROUPnoverb     -0.15370    0.24082  -0.638 0.523338
VERB_GROUPposture     0.06270    0.23049   0.272 0.785581
TR_WCpronoun          0.63322    0.20578   3.077 0.002090 **
TR_WCVP               0.54636    0.20992   2.603 0.009251 **
WO_LMtr_lm            0.32789    0.16261   2.016 0.043759 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1247.7  on 899  degrees of freedom
Residual deviance: 1047.2  on 889  degrees of freedom
AIC: 1069.2

Number of Fisher Scoring iterations: 4
```

Figure 4. Output for binary logistic regression model (multivariate corpus study)

The upper part of the output in Figure 4 specifies the model formula and the function used in producing the model in R. The next part of the output provides an overview by means of quartiles of the distribution of the deviance residuals – the difference between the observed and expected values. Since for (binary) logistic regression, the deviance residuals are expressed in logits, they need not follow a normal distribution. The following part of the output lists the estimates of the coefficients, which also pertain to logits, except for the variable

LM_LENGTHSYLLOG which is a numeric measure. The coefficient for LM_COMP (morphological complexity of the Landmark) which expresses the contrast between compound (the reference level of this variable which is not given in the model output) and simple Landmarks is, for example, positive (1.16). Positive coefficients indicate that the probability of the *peal*-construction increases. Hence, with simple Landmarks the probability of the *peal*-construction is higher. This result is confirmed by the univariate analysis, where it was demonstrated that compounds occur with the adessive construction and simple lexemes with *peal*. A negative coefficient, on the other hand, indicates that the probability of the adessive construction increases. For example, the coefficient for the variable LM_MOBILITY (mobility of the Landmark) which expresses the contrast between mobile (the reference level of this variable which is not given in the model output) and static Landmarks, is negative (-0.89). With static Landmarks the probability of the adessive construction is higher. This result again confirms the univariate results according to which mobile Landmarks prefer the *peal*-construction and static Landmarks the adessive construction.

Each estimated coefficient is accompanied by its estimated standard error, a Z-score, and the associated *p*-value. The line beneath the table of coefficients indicates that the dispersion parameter for the binominal family is taken to be 1. This is to remind that the variance of a binominal variable depends entirely on the mean, and that the model assumed that this property characterises our data. The next two lines in the summary provide the information necessary to check this assumption. The null deviance is the deviance one would get with a model with only an intercept. The difference between the null deviance and the residual deviance approximately follows a chi-squared distribution with, as degrees of freedom, the difference between the degrees of freedom of the two deviances. In R, this can be calculated using the following command for the values given in Figure 4: `1 - pchisq(1247.7 - 1047.2, 899 - 889)`. R returns "0" and this very small *p*-value shows that the model in Figure 4 has explanatory value. The final line of the output mentions the number of scoring iterations – 4 in the model in Figure 4. The algorithm for estimating the coefficients of a general linear model is iterative – it starts with an initial guess at the coefficients, and refines this guess in subsequent iterations until the guesses become sufficiently stable.

Before the actual results of the binary logistic regression modelling presented in the output in Figure 4 can be looked at, the issue of model criticism needs to be addressed. Although the small *p*-values in Figure 4 indicate that the specific levels of the variables in the model are significant[65], it may be the case that the model does not fit the data well enough or there may be other issues at

[65] It is not necessary for each level of the variable to be significant in order for the variable to be included in the model. Once at least one level of the variable (e.g. existence verbs) is significant, the variable needs to be included in the model. It does not matter that the other levels – motion verbs, no verb, and posture verbs (action verbs being the reference level here) – are not significant.

stake. The most important statistics provided by R that assess the goodness of fit of the model presented in Figure 4 are specified in Table 60.

Table 60. Goodness-of-fit statistics for binary logistic regression model (multivariate corpus study)

Statistic	Model 1
model L.R.	200.48 (df = 10)
generalised R^2	0.266
C	0.761
Somers' D_{xy}	0.522

Model L.R. stands for model likelihood chi-squared, the difference between the Null Deviance and the Residual Deviance that is encountered above in Figure 4. The degrees of freedom associated with it are given in the brackets. The remaining statistics address the predictive ability of the model. For normal regression models, the R^2 measure provides insight into how accurate the predictions of the model are. The problem with a dichotomous response variable like type of locative construction (adessive vs. *peal*) is that the model produces estimates of the probability that the construction is either the adessive or *peal*, whereas the observations simply state whether the construction is either the adessive or *peal*. Hence, for logistic regression R^2 is not as straightforward a measure as for normal regression. The measure named C is an index of concordance between the predicted probability and the observed response. When C takes the value of 0.5, the predictions are random, when it is 1, prediction is perfect. Baayen (2008: 204) reports that a value above 0.8 indicates that the model has real predictive capacity. The C value for model 1 is 0.761 – just below the critical 0.8. However, since 0.8 is given only as a suggestion for a critical value, 0.761 is considered a good enough value to conclude that model 1 has at least some predictive capacity. Somers' D_{xy} is a measure related to C. Somers' D_{xy} is a rank correlation between predicted probabilities and observed responses; it ranges between 0 (randomness) and 1 (perfect prediction). For model 1 Somers' D_{xy} is 0.522 which indicates some predictiveness. The model was also checked for collinearity by calculating the variance inflation factors; overdispersion does not appear to be a serious issue either.

In summary, the model presented in Figure 4 fits the data, but its predictive power is relatively weak. However, there is enough confidence in the model to inspect the results. The estimated coefficients and the associated measures given in Figure 4 indicate that with other predictors held constant, we would predict the likelihood of adessive usage to increase going from mobile Landmarks to static Landmarks. The likelihood of *peal*-constructions would be predicted to increase going from compound Landmarks to simple Landmarks, from action verbs to existence verbs, from nouns as Trajector phrases to pronouns and verbs as Trajector phrases, and from the word order 'lm_tr' to 'tr_lm'. These results mirror the univariate results reported in section 4.1.

4.4 Results of Distinctive Collexeme Analysis

The second corpus study takes a semantic perspective on the research question and looks at whether the two constructions exhibit any frequency preferences for certain type of lexemes. The first, multivariate corpus study examined the semantic variable 'type of Landmark' as well, but the dataset included only 900 extractions which are deemed insufficient to study lexical effects. Both the univariate and multivariate results of the first corpus study indicate that the semantic properties of the Landmark (particularly type and mobility) are strongly associated with the type of locative construction. Distinctive collexeme analysis (Stefanowitsch and Gries 2003, 2005; Gries and Stefanowitsch 2004a, 2004b; see section 3.1.2 above for an overview of this method) is used to find out how the two Estonian locative constructions differ from each other with respect to the lexemes they associate with. Different collocational preferences are taken to reflect the semantic differences between the constructions. The data, as described in section 3.1.2, comes from the fiction sub-corpus of the Balanced Corpus of Estonian (2008; size 5 million words). The calculation is based on four frequencies: the lemma frequency of the collexeme in construction A, the lemma frequency of the collexeme in construction B, the frequencies of construction A and B (Gries and Stefanowitsch 2004b: 102). These frequency tables are then submitted to the Fisher exact test. The analysis is run by using the open source software R together with a script written by Gries (2007a). The output of a distinctive collexeme analysis is a pair of lists, which rank collocating items of the two constructions according to their collostructional strength.

The final list of lexemes included 182 different nouns. The full frequency table with the 182 nouns (of which Table 61 represents a fraction) were submitted to the analysis in R. The words were then ranked according to the collostructional strength and the results are presented in Table 61. Table 61 lists the 14 distinctive collexemes for the adessive case construction and the 33 distinctive collexemes for the adposition *peal* 'on'. The numbers in the brackets indicate the ratio between the adessive and *peal*-constructions. For example there were 580 adessive constructions vs. 2 *peal*-constructions with the lexeme *tänav* 'street' and 22 *peal*-constructions vs. 38 adessive constructions with the lexeme *hing* 'soul' in the dataset. The nouns are listed according to the measure of collostructional strength; collostructional strength values that are larger than 1.3 indicate that a lexeme is distinct at the significance level of $p < .05$ (Hilpert 2006: 245). The more distinct a lexeme is, the higher its numerical value of collostructional strength (*coll. str.*). All of the lexemes listed in Table 61 have a collocational strength higher than 1.3 and are all therefore judged to be significantly distinctive.

Table 61. Collexemes distinguishing between the adessive case and the adposition *peal* 'on'

ADESSIVE (N=7979)		PEAL 'ON' (N=703)	
Collexeme	*Coll. str.*	*Collexeme*	*Coll. str.*
tänav 'street' (580:2)	18.93	*hing* 'soul' (22:38)	9.29
turg 'market' (437:1)	14.88	*sohva* 'couch' (12:20)	5.45
nägu 'face' (300:4)	6.90	*hoov* 'yard' (9:16)	4.05
saar 'island' (229:2)	6.21	*kelk* 'sledge' (5:6)	2.98
lava 'stage' (283:5)	5.66	*kapp* 'wardrobe' (4:4)	2.64
kallas 'shore' (229:5)	4.06	*peegel* 'mirror' (4:4)	2.64
ekraan 'screen' (121:1)	3.43	*näpp* 'finger' (5:8)	2.59
õlg 'shoulder' (135:2)	3.12	*kõht* 'stomach' (8:24)	2.49
plats 'square' (185:5)	2.84	*pilv* 'cloud' (5:9)	2.43
põld 'field' (118:2)	2.59	*pilt* 'picture' (35:236)	2.43
katus 'roof' (90:1)	2.40	*laud* 'table' (47:345)	2.41
pind 'surface' (149:4)	2.37	*voodi* 'bed' (14:66)	2.34
laev 'ship' (100:2)	2.06	*ämber* 'bucket' (3:2)	2.33
diivan 'couch' (136:4)	2.03	*rahakott* 'wallet' (3:2)	2.33
		ratas 'wheel' (7:20)	2.33
		ots 'end' (4:6)	2.22
		redel 'ladder' (4:6)	2.22
		aken 'window' (17:93)	2.16
		maamuna 'globe' (5:12)	2.03
		nurk 'corner' (14:78)	1.81
		õu 'yard' (16:95)	1.78
		muld 'soil' (4:9)	1.77
		süda 'heart' (7:27)	1.76
		jää 'ice' (8:34)	1.75
		veesoon 'vein of water' (2:1)	1.73
		kivi 'stone' (11:57)	1.70
		lina 'sheet' (4:10)	1.66
		pesa 'nest' (3:6)	1.51
		raamat 'book' (3:6)	1.51
		tasku 'pocket' (2:2)	1.45
		postament 'pedestal' (3:7)	1.38
		riie 'cloth' (3:7)	1.38
		põlv 'knee' (12:75)	1.33

In interpreting the results, it has to be born in mind that one of the major limitations of distinctive collexeme analysis is that it does not take into account the overall corpus frequencies of the lexemes that occur in the two constructions – it merely highlights differences and does not characterise the two constructions per se (Hilpert 2006: 250). It is therefore very likely that the most typical lexeme for the adessive construction does not occur at all with the *peal*-construction and vice versa. Distinctive collexeme analysis does not consider this issue since it only looks at words that are used in both constructions. In order to provide a more adequate characteristic of the two constructions, two

separate collexeme analyses would have to be run. This is not undertaken in the present study, because the focus is on looking at the two constructions simultaneously, i.e. the main interest is precisely on these lexemes that are jointly used by both of the constructions. Still, it is important to bear this limitation in mind when looking at the results and the discussions that follow.

Overall, the results support a semantic approach, in that there are collexemes that clearly distinguish between the adessive case construction and the adpositional construction with *peal* 'on'. Furthermore, the specific claims made in previous research on the alternation between cases and adpositions (e.g. Bartens 1978, Ojutkangas 2008) and the findings of multivariate corpus study are largely confirmed – the adessive case is used with nouns denoting places and the adposition *peal* 'on' with nouns denoting smaller, manipulable things. For the adessive case, the most distinctive collexeme is *tänav* 'street' – a noun that clearly denotes a static, large place and not a small, manipulable thing. Together with *tänav* 'street', 9 other lexemes (*turg* 'market', *saar* 'island', *lava* 'stage', *kallas* 'shore', *plats* 'square', *põld* 'field', *katus* 'roof', *pind* 'surface', *laev* 'ship') out of the total of 14 lexemes that are distinctive for the adessive case, denote a place. Out of 14 distinctive collexemes for the adessive case there are only 4 nouns that do not denote a place – 2 nouns refer to smaller and manipulable things (*ekraan* 'screen' and *diivan* 'couch') and 2 nouns refer to body parts (*nägu* 'face' and *õlg* 'shoulder').

For the adpositional construction with *peal* 'on' the most distinctive collexeme is *hing* 'soul'. Although this lexeme does not denote a small, manipulable entity, it does confirm another result found in multivariate corpus study. As we saw above in section 4.1.1, there was a significant difference between the two constructions as to the type of relationship between a Trajector and Landmark. The results of multivariate corpus study demonstrate that the adposition *peal* is significantly more often used with abstract (inc. figurative) relations than the adessive construction. The lexeme *hing* 'soul' is one such frequent Landmark found in the data that expresses an abstract relation between a Trajector and Landmark and that is used with the adpositional construction rather than with the case construction. Another Landmark that is used in abstract relations is *süda* 'heart', which according to the results of the analysis presented in Table 61 is a distinctive collexeme for the *peal*-construction.

In addition, there is also support for the prediction that the adpositional construction is used with small, manipulable entities – out of the total of 33 collexemes strongly distinctive for the *peal*-construction, as many as 19 denote things: *sohva* 'couch', *kelk* 'sledge', *kapp* 'wardrobe', *peegel* 'mirror', *pilt* 'picture', *laud* 'table', *voodi* 'bed', *ämber* 'bucket', *rahakott* 'wallet', *ratas* 'wheel', *ots* 'end', *redel* 'ladder', *kivi* 'stone', *lina* 'sheet', *raamat* 'book', *tasku* 'pocket', *postament* 'pedestal', *riie* 'cloth', *pesa* 'nest'. At the same time, 8 collexemes (*hoov* 'yard', *aken* 'window', *maamuna* 'globe', *nurk* 'corner', *õu* 'yard', *muld* 'soil', *jää* 'ice', *veesoon* 'vein of water') denote a place which characterises the adessive construction. However, it should be noted that other variables besides semantics may play a role here. As we saw from the results of

multivariate corpus study above (sections 4.1.1 and 4.1.2), the length of the Landmark phrase was a very significant variable that has a strong significant effect on the choice between the two constructions. It is therefore important to note that out of the 8 lexemes that denote a place, but that are contrary to expectations distinctive for the adpositional construction and not the case construction, 5 are monosyllabic in the nominative form – *hoov* 'yard', *nurk* 'corner', *õu* 'yard', *muld* 'soil', and *jää* 'ice'. Furthermore, while the collexemes *maamuna* 'globe' and *veesoon* 'vein of water' that are distinctive for the *peal*-construction are clearly not small things, but denote larger places and are also relatively long words, both of these have lexicalized from nouns that in fact denote small, manipulable things. The *muna* component in *maamuna* literally refers to an egg and *soon* in *veesoon* refers to a vein.

A similar morphosyntactic explanation can be given for the two nouns denoting things that are distinctive for the adessive construction – *ekraan* 'screen' and *diivan* 'couch' – both are relatively long words and longer Landmark phrases were found to prefer the adessive construction. It is also important to note that the two lists in Table 61 both contain a noun denoting a couch – *diivan* in the list of distinctive collexemes of the adessive construction and *sohva* in the list of distinctive collexemes of the *peal*-construction. The reasons for this apparent anomaly can be either related to length – *diivan* is longer than *sohva* when we count length in orthographic characters – or to the general frequency of these two nouns. A quick query demonstrates that the lexeme *diivan* is much more frequent (384 tokens) than the lexeme *sohva* (84 lexemes) in the 5-million word fiction sub-corpus of BCE.

As we saw above in Table 23 in section 4.1.1 which presents the results for the variable 'type of Landmark' in multivariate corpus study, with body parts as Landmarks the two constructions were equally frequent. This result is confirmed by the distinctive collexeme analysis as both lists in Table 61 include body parts: *nägu* 'face' and *õlg* 'shoulder' are distinctive for the adessive construction and *näpp* 'finger', *kõht* 'stomach', and *põlv* 'knee' for the *peal*-construction. It is difficult to deduct why specifically these body part nouns are distinctive for the respective constructions. It may be hypothesised, at least for the body part *nägu* 'face', that the reason why it is preferred by the adessive construction has to do with the fact that the adessive form of this body part has grammaticalized into an adpositional usage *näol* ' in the form of' as exemplified in 65:

(65) *See , mis* ***kuulujuttude*** ***näol*** *juba pikka*
this.SG.NOM that gossip.PL.GEN face.SG.ADE already long.SG.PRT
aega rahva seas ringi liikus , hakkas
time.SG.PRT people.SG.GEN among around move-PRS.3SG begin-PST.3SG
lõpuks tõeks saama . <BCE: fiction>
finally truth.SG.ESS become-INF
'That what had been moving around among people already for a long time **in the form of gossip** (lit. on the face of gossip) was finally becoming true.'

Although such instances of the use of *näol* were excluded from the present analysis – *peal*-construction is not an interchangeable alternative in this instance – it may be the case that precisely due to the fact that this lexeme has already grammaticalized in Estonian (i.e. it has a relatively high frequency of usage compared to those body part terms that have not grammaticalized into adpositions), language speakers prefer the case construction with this body part also when they feel the need to express the situation where something is physically on somebody's face, as in example 66 below.

(66) *Öösel toibudes oli mu keel*
night.SG.ADE recover-PRS.PTCP be-PST.3SG my tongue.SG.NOM
*verine , **näol** aga mitu tursunud sinikat.*
bloody.SG.NOM face.SG.ADE but many swollen bruise.SG.PRT
<BCE: fiction>
'In the night when I was recovering, my tongue was bloody, and there were many swollen bruises **on my face**.'

As pointed out by Gries and Stefanowitsch (2004b: 107), it is also useful to look at nouns that are *not* distinctive for either construction, i.e. nouns that readily occur in both constructions. Out of the total of 182 nouns used in this analysis, as many as 135 were not distinctive; the numbers of these lexemes are given in Table 62 in eight groups according to the type of Landmark.

Table 62. Types of lexemes not distinctive for either the adessive or *peal*-construction

Type of Landmark	Number of lexemes	Proportion
abstract	13	10%
body part	18	13%
building	6	4%
clothing	11	11%
furniture	14	14%
place	35	26%
thing	33	24%
vehicle	5	4%
Total	135	100%

It is somewhat surprising that a large number of lexemes are not distinctive for either the adessive case or the adposition *peal*, even though we saw from the results of multivariate corpus study that there is a significant effect for the variable 'type of Landmark' as pertains to the choice between these two constructions. Although there were 13 lexemes denoting places and 14 lexemes denoting things that were distinctive, as many as 35 place-lexemes and 33 thing-lexemes were not distinctive. However, there is also converging evidence as both the results of multivariate corpus study (Table 23 in section 4.1.1 above) and the distinctive collexeme analysis (Table 62) show that Landmarks denoting body parts, furniture and clothing occur equally frequently with both constructions. Even if 5 lexemes denoting body parts, 7 furniture and 1 clothing were distinctive, the number of these types of lexemes that are not distinctive is considerably larger – 18, 14, and 11 respectively.

The results of the distinctive collexeme analysis point towards two possible conclusions. The relatively large number of lexemes that are not distinctive for either of the two constructions seems to indicate that the two constructions are in fact near-synonyms – although overall the adessive case construction is over 10 times as frequent as the *peal*-construction in expressing the support-relation with these 182 lexemes, in many cases the two constructions are both possible. The other possible interpretation of these results is that semantics is not, in fact, the biggest piece of puzzle and that there are other variables that are more important than the semantics of the Landmark. As we saw from the multivariate results of the first corpus study, this seems to be the case – logistic regression models show that the most significant variable that contributes to the choice between the adessive and the adposition *peal* in written present-day Estonian is the morphosyntactic variable length. On the other hand, the fact that there are distinctive collexemes for the two constructions and that the semantic variables type and mobility of Landmark had a significant effect on the choice does not allow us to disregard semantics entirely. It is clear that two separate collexeme analyses are required in order to provide more reliable results as to the constructional profiles of these two constructions. However, the aim here is to take a purely contrastive look at the phenomenon.

It is pointed out in the literature that distinctive collexeme analysis has shown that many alternations are much more restricted than has been previously assumed (Gries and Stefanowitsch 2004b: 97). From the perspective of the alternation between Estonian locative cases and adpositions the conclusion to be drawn, instead, is that it seems this specific alternation is much less restricted than the theories of Cognitive Grammar (Langacker 1987, 2008) and Construction Grammar (Goldberg 1995, 2006) would lead us to believe. Although there do appear to be subtle semantic differences between the two constructions in that case constructions tend to prefer larger entities like places and buildings, and the adpositional construction abstract entities and small, manipulable things, these differences are comparatively small. This points to the more general conclusion that even though the two constructions are not full synonyms, they do appear to be near-synonyms in written present-day Estonian.

4.5 Summary of corpus results

For the multivariate corpus study, 900 sentences (450 with the adessive and 450 with the *peal*-construction) from present-day written Estonian (both fiction and newspaper texts) are analysed for 20 different variables – 11 semantic and 9 morphosyntactic. It is predicted that the two Estonian locative constructions behave differently vis-à-vis these variables. The univariate results show that 9 (five morphosyntactic and four semantic) out of these 20 are significantly associated with the type of construction. However, given the fairly large dataset, there are only three variables with a relatively strong effect size. All are **morphosyntactic variables**: length of the Landmark phrase in syllables and words, and morphological complexity. The fact that there are not very many variables that have a large effect size is maybe not surprising given the complex and multivariate nature of the phenomenon. The three strongest morphosyntactic variables and the two other (word class of Trajector and Landmark) reflect in general the basic trend – the synthetic adessive construction tends to occur with longer and more complex Landmark phrases, while the analytic *peal*-construction is preferred with shorter and less complex (e.g. pronominal) Landmark phrases. This seems to indicate that the principle of language economy is at play. Language users prefer the shorter and more concise construction, i.e. the case construction, in environments that are otherwise cognitively more complex. The more specific and explicit adpositional construction is preferred in environments that are less explicit themselves and thus require the locative construction to carry extra specificity. The morphosyntactic variables of clause type, the syntactic function of the Landmark, case form of the Trajector, position of the Landmark phrase, and the relative position between Trajector and Landmark are not found to play a role in the dataset studied for this dissertation.

The strongest **semantic variable** according to the univariate results is mobility of the Landmark, followed by relative size of the Trajector in relation to the Landmark, type of Landmark, and verb group. In general, it can be concluded that the adessive case is used with larger static places. The *peal*-construction is preferred with smaller, movable things and with verbs of existence. The different semantic variables pertaining to Trajectors (type, mobility, number, and animacy of the Trajector) along with the variables animacy and number of Landmark, and the type of relation between Trajector and Landmark showed no significant effect in the corpus results.

The **multivariate results** of the first corpus study confirm the relevance and significance of the following variables: length of the Landmark in syllables, morphological complexity of the Landmark, mobility of the Landmark, verb, word class of Trajector, and the relative position between Trajector and Landmark. This is not to say that the other variables found to have an effect in the univariate analysis are completely irrelevant. It is just that for the present data, the best model that could be fitted included the combination of these variables. In general, the multivariate results confirm the univariate results:

longer and more complex static locations predict the adessive construction, while shorter and more simple environments with verbs of existence and mobile Landmarks predict the adpositional construction. However, it should be kept in mind that no particularly strong model could be fitted to the data. The goodness-of-fit statistics and plots used for model criticism allow for some confidence in the strength and predictive power of the model, but the regression models reported for the various alternation phenomena in English exhibit a much better fit to the data and more powerful predictions. One of the conclusions to be drawn from the fact that, in theory, the model could have more explanatory power is that there may be other variables not included in the present analysis that play an important role in determining the use of these two constructions. For instance, all of the discourse-functional variables, such as topic, register, preceding and subsequent mention of the adessive or adpositional construction, and variables like idiolect and dialect are absent from the present analysis. Nevertheless, both the univariate and multivariate analyses of the present data also systematically indicate that there are significant (if comparatively moderate) differences between the Estonian adessive and the adposition *peal* 'on'.

The **second corpus study** takes a purely semantic look at the two constructions and uses the technique of distinctive collexeme analysis. The aim of this study is to pinpoint specific lexemes that distinguish between the two constructions. In essence, the study is used to find converging evidence for the semantic variable 'type of Landmark'. Since the data collected for this study included only lexemes that appear at least once with each construction, care should be taken in interpreting the results. The list of possible distinctive lexemes arrived at in the present dissertation only includes lexemes that distinguish between the two constructions. These lexemes may not be the most characteristic lexemes for the two constructions in general, since the lexemes that only occur with one type of construction are not included in the study. From the perspective of the research question of the dissertation the study fulfils its aim – to look at only these lexemes that are possible in both constructions. Nothing is claimed about the general characteristics of these two constructions in isolation. In general, the results of the distinctive collexeme analysis confirm the results of the multivariate corpus study. Out of the total of 182 lexemes that are found to be used with both the adessive and *peal*-construction in the 5-million written corpus of Estonian, 14 are distinctive for the adessive and 33 for *peal*. The majority of these 14 distinctive collexemes of the adessive construction denote large static places (e.g. *tänav* 'street', *turg* 'market', *saar* 'island'); and the majority of the 33 distinctive collexemes of the *peal*-construction denote smaller, manipulable things (e.g. *sohva* 'couch', *kelk* 'sledge', *kapp* 'wardrobe').

The next chapter looks at the experimental results where the relevance of the semantic variable 'type of Landmark' and two morphosyntactic variables – length and relative position of Trajector and Landmark – in the alternation between the adessive and *peal*-construction is assessed in two rating tasks. The

multivariate corpus study and the distinctive collexeme analysis point towards the conclusion that the semantic properties of the Landmarks play a significant role in the choice between the adessive case and the adposition *peal*. It was decided to examine if and how production maps onto comprehension and to see if language users are aware of the conceptual distinctions that these two locative constructions exhibit. This is the aim of the first experiment – the picture rating task. The binary logistic regression model fitted to the multivariate corpus data also showed that 'length of the Landmark phrase' is one of the strongest predictors for the Estonian locative alternation. The second experiment, sentence rating task, examines this morphosyntactic variable. The two experimental tasks also assess the relevance of word order. Although the univariate results of the multivariate corpus analysis show no statistically significant association between word order and the two locative constructions, the binary logistic regression analysis indicates that once the variable 'relative position between Trajector and Landmark' is factored in together with length, complexity, and mobility of the Landmark, verb group, and word class of Trajector, it becomes somewhat significant. It is also my own native-speaker intuition that word order may play a role and it was therefore decided to explore this issue further.

5. RESULTS: EXPERIMENTAL STUDIES

This chapter reports the results of the two experimental tasks: the picture rating task (section 5.1, with the discussion of results in 5.2) and the sentence rating task (section 5.3, with the discussion of results in 5.4). The picture rating task looks at the semantic variable type of Landmark and the morphosyntactic variable word order. These variables were selected to validate the claim that, first of all, the case construction is used with larger, static places and the adpositional construction with smaller, mobile things (a result obtained with both the multivariate corpus study and the distinctive collexeme analysis), and secondly that the adessive construction is preferred when the Landmark precedes the Trajector and the adpositional construction when it follows the Trajector. Since the two constructions express spatial relations, it was felt necessary to provide visual stimuli. The sentence rating task looks at two morphosyntactic variables – length of the Landmark phrase and the relative position of Trajector and Landmark. The reason why specifically these variables were selected has to do with converging evidence. The multivariate corpus study shows that length of the Landmark phrase is one of the strongest predictors – the longer the Landmark phrase, the more probable it is that the adessive construction is used. It was decided to validate this result in a sentence rating task. The picture rating task provided conflicting results as pertains to the variable relative position of Trajector and Landmark. The multivariate corpus results indicate that there is a slight preference for the adpositional construction to be used when the Landmark follows the Trajector – this is also my intuition as a native speaker of Estonian. However, the picture rating task leads to a different conclusion – in the experimental conditions for the picture rating task the adessive received higher ratings when the Landmark followed the Trajector than the corresponding sentences with the adpositional construction. It was therefore decided to include this variable in the sentence rating task as well.

The presentation of the results for both rating tasks follows a similar logic: first, the descriptive analysis of the datasets is given (mainly graphical examinations of the data), followed by the inferential part which presents the results of the statistical methods that test the validity of the hypotheses. Although the most fool-proof way to approach these tasks is to consider the response variable (the rating provided by the participants) as a measurement on an ordinal (i.e. categorical) scale and not on an interval (i.e. continuous) scale, the discussion of results also includes both types of methods. The reason for this is to see if there are any significant differences between the results. For a researcher not overly familiar with different regression modelling techniques, analysis of variance seems easier to use. It is the gold standard in much of psycholinguistic research and the results are easier to interpret than the results of ordinal logistic regression. At the same time, it is important to remember that I regard the ratings in essence as measurements on ordinal scale, but as the discussion in section 3.2 demonstrates, there are arguments in the literature for *treating* the ratings as measurements on interval scale. The analysis of the data

is carried out using R (version 2.10.1, R development core team 2009). On occasion (e.g. for ordinal logistic regression), IBM SPSS Statistics 19 software is used.

5.1 Results of picture rating task

Before presenting the actual results, a few key aspects about the design of the picture rating task are given here as a reminder; the full details are given above, in section 3.2.1. The picture rating task is a 5 (type of Landmark: body part, building, furniture, place, thing) x 2 (relative word order between the Landmark and the Trajector: lm1, lm2) x 2 (type of locative construction: adessive, *peal* 'on') factorial design. For each of the 20 experimental conditions, 4 different lexicalizations of the Landmark and Trajector were used. The response variable is a rating given on a scale from 1 ('not at all probable') to 5 ('very probable') by the participants who were asked to rate how probable it is that they themselves would use the sentence given to describe the picture shown together with the sentence. The questionnaire was distributed electronically using the eFormular tool and 73 participants participated in this task. The picture rating task tested the influence of the semantic variable type of Landmark and the morphosyntactic variable word order. Both of these variables were included in the multivariate corpus study. Both the univariate and multivariate results of the first corpus study and the results of the distinctive collexeme analysis show that the type of Landmark is a significant variable in predicting the choice between the adessive and *peal*-construction. The univariate results of the multivariate corpus study did not show a significant association between the relative word order between the Landmark and Trajector, but the multivariate results indicate that once other variables like length, complexity, and mobility of the Landmark, word class of Trajector, and verb group are factored in, the relative word order between the Landmark and Trajector becomes significant. In order to verify that there is indeed a significant association between the type of Landmark and construction type and that language users are sensitive to it, and to check if and how word order influences the choice, the picture rating task was designed. The results of this task are presented below.

5.1.1 Description of the results

The most important aspect of the data set for this experiment is that the data is heavily skewed. If we look at Figure 5 which presents the overall frequency of the data (averaged over all observations), we see that the data is not normally distributed and is heavily skewed towards the right, i.e. towards higher ratings. As many as 1,228 observations out of a total of 2,920 fall into the category of rating 5 (42%); 5 is also the median rating averaged over all of the data points. Still, as we saw in section 3.3, according to the Central Limit Theorem, for a sample size of 75 participants, the sampling distribution of the sample mean is approximately a normal distribution (Agresti and Finlay 2009: 93).

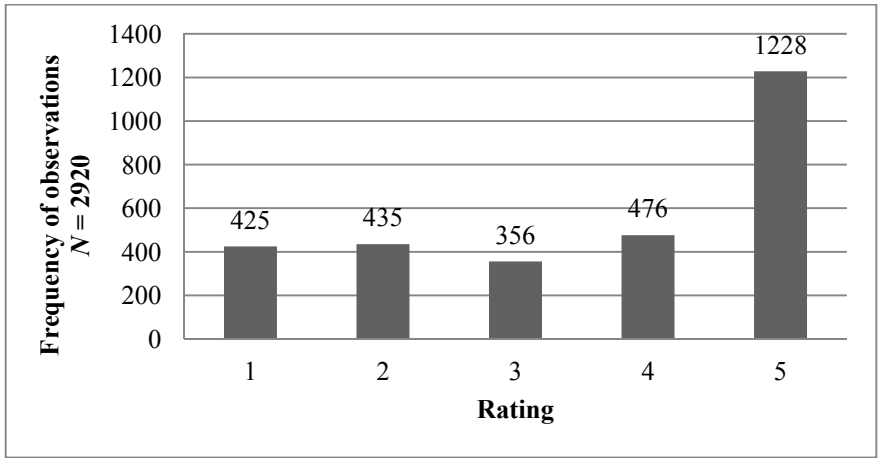


Figure 5. Frequency of ratings averaged over all observations in picture rating task

The measure of central tendency for ordinal data is the median – the value separating the higher half of the data sample from the lower half. Another measure of central tendency for ordinal data is the mode – the value that occurs most frequently in a data set. The central tendency for interval data is the mean, but it is pointed out in statistics handbooks that for skewed data, the median is more appropriate. Table 63 provides the medians, modes, means and standard deviations for all of the 20 conditions (arranged in albhapetical order).

Table 63 already highlights the fact that whether we take the dependent variable as continuous or categorical, one may end up with different conclusions about the outcome of the experiment. Since the data is heavily skewed, it is no wonder that the mode is not very informative in this case – for almost all of the 20 conditions it is 5. Only three conditions depart from this general trend – adessive sentences with things where the Landmark comes first (ade_thing_lm1), peal sentences with places where the Landmark comes first (peal_place_lm1), and peal sentences with places where the Landmark comes second (peal_place_lm2). In the latter condition, there are actually two modes – 1 and 5. This situation highlights a problem researchers may encounter when they take the mode as the central tendency. Since mode is the value that occurs most frequently, there might be two or more values in the data set that occur with the same frequency. However, it does serve as a good example to demonstrate how important information might be lost if we just look at the mean. The mean for the condition 'peal_place_lm2' is 3.0 (SD = 1.5) and the median is also 3.

Table 63. Median, mode and mean ratings for all experimental conditions in picture rating task[66]

Condition	Med.	Mode	Mean	SD	Condition	Med.	Mode	Mean	SD
ade_bodypart_lm1	3	5	3.0	1.6	peal_bodypart_lm1	4	5	3.6	1.5
ade_bodypart_lm2	4	5	3.6	1.5	peal_bodypart_lm2	4	5	3.6	1.4
ade_building_lm1	5	5	3.7	1.5	peal_building_lm1	3	5	3.2	1.5
ade_building_lm2	5	5	4.1	1.2	peal_building_lm2	4	5	3.3	1.5
ade_furniture_lm1	4	5	3.3	1.5	peal_furniture_lm1	4	5	3.9	1.3
ade_furniture_lm2	4.5	5	3.8	1.4	peal_furniture_lm2	5	5	4.2	1.2
ade_place_lm1	4	5	3.8	1.4	peal_place_lm1	3	1	2.9	1.5
ade_place_lm2	5	5	4.0	1.4	peal_place_lm2	3	1; 5	3.0	1.5
ade_thing_lm1	2	1	2.6	1.5	peal_thing_lm1	5	5	3.8	1.4
ade_thing_lm2	4	5	3.6	1.5	peal_thing_lm2	5	5	4.0	1.2

[66] The code to interpreting the conditions is the following: 'ade' represents adessive sentences, 'lm1' sentences with the Landmark phrase preceding the Trajector phrase, 'lm2' sentences with the Landmark phrase following the Trajector phrase. The condition 'ade_bodypart_lm1' therefore represents sentences where the locative construction is adessive, type of Landmark is bodypart, and word order is such that the Landmark comes first.

For clarity and a better overview of the data, information given in Table 63 is presented visually in Figure 6, where the conditions are arranged not in alphabetical order as in Table 63, but from lowest to highest rating. Because the mode in this case is not very informative and the use of the median is less controversial than the use of the mean, the median was used for producing Figure 6.

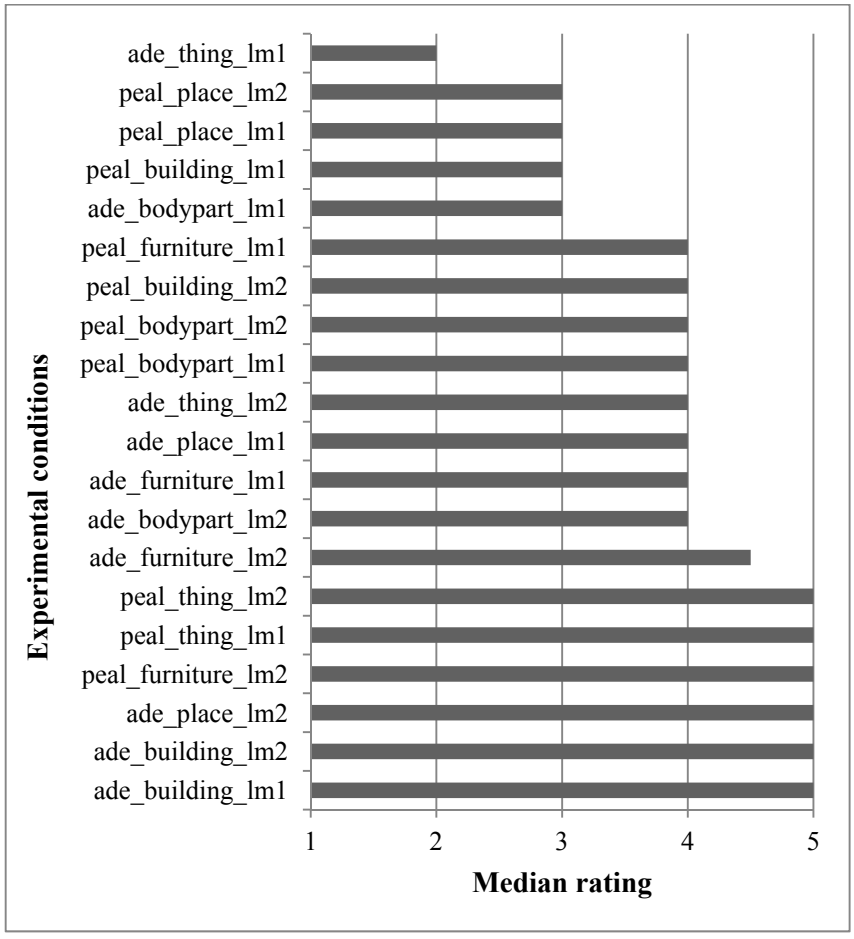


Figure 6. Median ratings for all experimental conditions in picture rating task

It can be seen from Table 63 and Figure 6 that there is no sharp contrast between sentences with the adessive case and the adposition *peal* 'on'; what we have instead is a continuum of ratings. There are both adessive and *peal*-sentence types clustering at the lower (sentences with the median rating of 2 or 3) as well as the higher end of the continuum (sentences with the median rating of 4 or 5). This is confirmed if we also look at Figure 7 which presents the

frequency counts for the two sentence groups according to the five rating categories – the data is equally skewed for both adessive and *peal*-sentences.

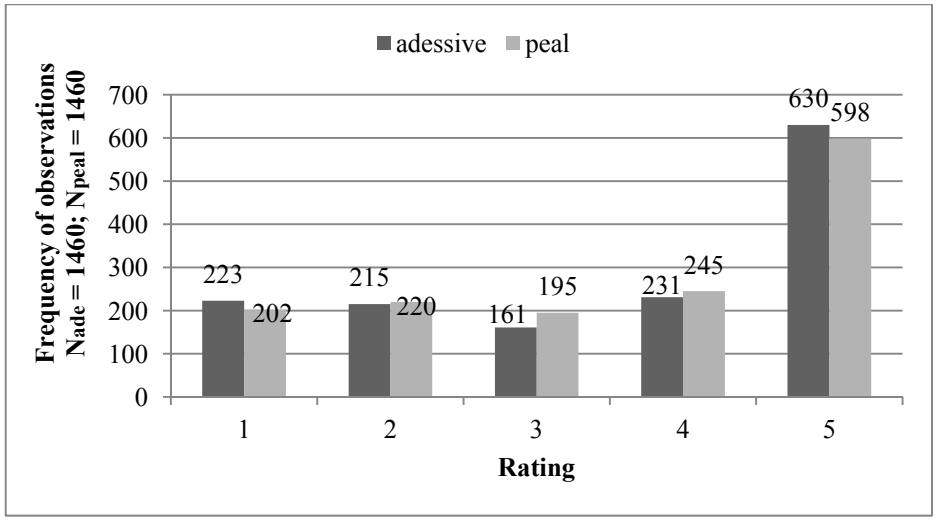


Figure 7. Frequency of ratings for adessive and *peal*-sentences averaged over all observations in picture rating task

It can also be seen from Figure 7 that the two distributions are very similar – they are both skewed towards higher ratings. Averaged over all observations, the rating for the adessive sentences and *peal*-sentences did not differ – the median was 4 and the mode 5 for both groups of sentences. The mean ratings show the same result – the mean acceptability rating for the adessive sentences and *peal*-sentences was more or less the same – 3.57 for the former (SD = 1.52) and 3.56 for the latter (SD = 1.48). These results are confirmed by both a Wilcoxon test (V = 538.5, $p = 0.99$) and a *t*-test for dependent samples ($t = 0.09$, $df = 72$, $p = 0.93$).

Following is an exploratory analysis of the data (mainly graphical examinations) with the inferential statistics presented in sections 5.1.7–5.1.10. The reason for including exploratory data analysis is to provide a more intuitive overview of the nature of the dataset. However, let it be said in anticipation that the different inferential statistical methods employed confirm a significant main effect for type of Landmark and significant interactions between type of construction and type of Landmark, and between type of construction and word order. The experiment is designed so that each factor is crossed with each level of the other factors; this allows inspecting the interactions between different explanatory variables.

5.1.2 Type of Construction x Type of Landmark

It is informative from the perspective of the research question to take a look at the participants' ratings for the sentences containing different types of Landmarks. Generalised over all observations, Table 64 provides the medians, modes, means and standard deviations for all of the five Landmark types. It can be seen that the medians for all five types are the same – 4. If we compare the means, then the mean rating for furniture is a little higher (mean = 3.79, SD = 1.39) than for the other four types of Landmark.

Table 64. Median, mode and mean ratings for the five types of Landmark in picture rating task

Landmark	Median	Mode	Mean	SD
bodypart	4	5	3.46	1.53
building	4	5	3.59	1.49
furniture	4	5	3.79	1.39
place	4	5	3.43	1.55
thing	4	5	3.53	1.52

It would be interesting to see if these figures differ when we look at the adessive and *peal*-sentences separately. Generalised over all observations, Table 65 provides the medians, modes, means and standard deviations for all of the five Landmark types with the two locative constructions. Table 65 points to some of the differences that become crucial in the statistical models described in second part of the analysis: adessive sentences with buildings and places as Landmarks received a higher average median and mean rating than the corresponding *peal*-sentences; *peal*-sentences with furniture and things as Landmarks received higher average median and mean ratings than the corresponding adessive sentences; there does not seem to be a difference between the two constructions types with bodyparts. This interaction is graphically presented in Figure 8.

Table 65. Median, mode and mean ratings for types of Landmarks according to the two locative constructions in picture rating task

Type of	adessive sentences				peal-sentences			
Landmark	Median	Mode	Mean	SD	Median	Mode	Mean	SD
bodypart	4	5	3.32	1.58	4	5	3.60	1.47
building	**5**	5	3.90	1.40	**3**	5	3.29	1.53
furniture	**4**	5	3.58	1.48	**5**	5	4.02	1.26
place	**5**	5	3.91	1.40	**3**	5	2.95	1.54
thing	**3**	5	3.14	1.59	**5**	5	3.93	1.33

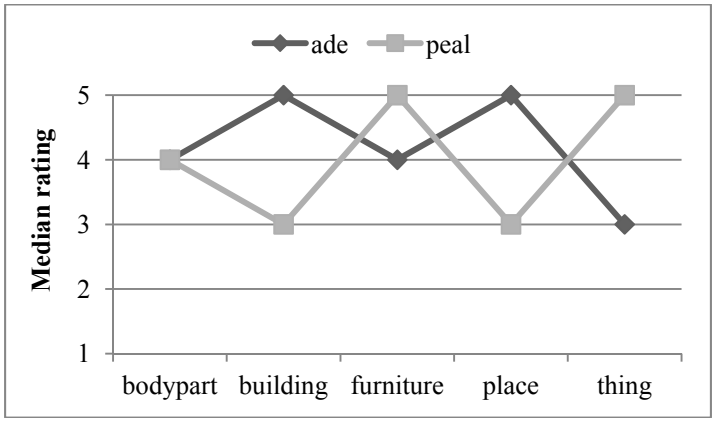


Figure 8. Interaction plot for type of construction (adessive vs. *peal*) and type of Landmark (median ratings) in picture rating task

5.1.3 Type of Construction x Word order

It is also informative to take a look at the participants' ratings for the sentences with different word order sequences – sentences where the landmark precedes the trajector (landmark comes first = "lm1") versus sentences where the landmark follows the trajector (landmark comes second = "lm2"). Generalised over all sentences, the median ratings, modes, means and standard deviations for both word order sequences are given in Table 66. The medians and modes are the same for both word order sequences, although the mean rating seems to indicate that the Landmark second word order is rated higher than the Landmark first word order.

Table 66. Median, mode and mean ratings for the two word order sequences in picture rating task

Word order	Median	Mode	Mean	SD
lm1	4	5	3.39	1.55
lm2	4	5	3.73	1.44

Furthermore, dividing the sentences into two groups according to the type of construction pinpoints at least one difference. Table 67 shows that with respect to the word order sequence where the landmark precedes the trajector ("lm1"), there is no difference (the average median rating for both the adessive and *peal*-sentences is still 4), but participants rated the adessive sentences where the landmark followed the trajector ("lm2") higher (the median was 5) compared to the corresponding *peal*-sentences (with a median of 4).

Table 67. Median, mode and mean ratings for the two word order sequences according to the two locative constructions in picture rating task

Word order	adessive sentences				peal-sentences			
	Median	Mode	Mean	SD	Median	Mode	Mean	SD
lm1	4	5	3.28	1.57	4	5	3.50	1.51
lm2	**5**	5	3.85	1.42	**4**	5	3.62	1.45

This interaction is illustrated in Figure 9.

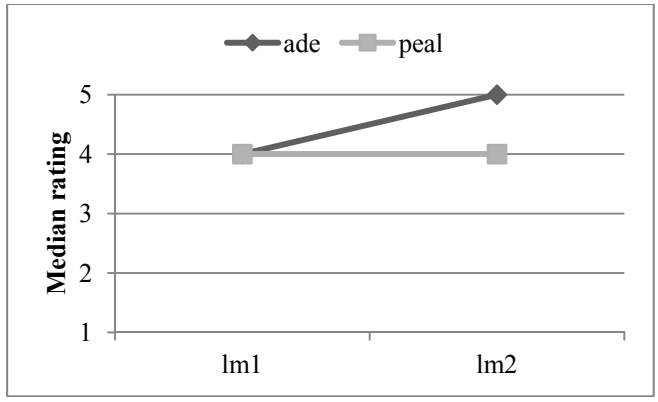


Figure 9. Interaction plot for type of construction (adessive vs. *peal*) and word order in picture rating task

This result is different from the results found for the multivariate corpus study, where the logistic regression model indicates that the "lm2" word order sequence predicts the adpositional construction. The reason why there is diverging evidence may have to do with the different set-up of the two studies. The logistic regression model fitted to the corpus data takes into account the other variables as well – length, complexity, and mobility of Landmark, word class of Trajector, and verb group. The relative position of the Landmark and Trajector is significant only when it is factored in simultaneously with the other variables. With the picture rating task, which only includes short Landmark phrases, the language users prefer the adessive construction to come at the end of the sentence.

5.1.4 Type of Landmark x Word order

In addition to the two interactions described above – the combination of the two construction types with the type of Landmark and word order – there is a third two-way interaction that can be looked at: type of Landmark and word order.

Table 68 provides the medians, modes, means and standard deviations for all of the five Landmark types with the two word order sequences.

Table 68. Median, mode and mean ratings for types of Landmarks according to the two word order sequences in picture rating task

Type of Landmark	lm1				lm2			
	Median	Mode	Mean	SD	Median	Mode	Mean	SD
bodypart	4	5	3.31	1.58	4	5	3.60	1.47
building	4	5	3.46	1.56	4	5	3.73	1.42
furniture	**4**	5	3.60	1.45	**5**	5	4.00	1.31
place	4	5	3.37	1.55	4	5	3.50	1.55
thing	**3**	5	3.23	1.57	**4**	5	3.84	1.40

Sentences with body parts, buildings and places show no difference between the median ratings in regard to whether the landmark precedes or follows the trajector (the median rating for all combinations is 4). However, there is a difference between the two word order sequences with sentences containing pieces of furniture and things. Participants rated the combination of furniture as landmark and thing as landmark following the trajector (condition "lm2") higher (median ratings 5 and 4 respectively) than the combination of furniture as landmark and thing as landmark preceding the trajector (condition "lm1", median ratings 4 and 3 respectively). This interaction is presented graphically in Figure 10.

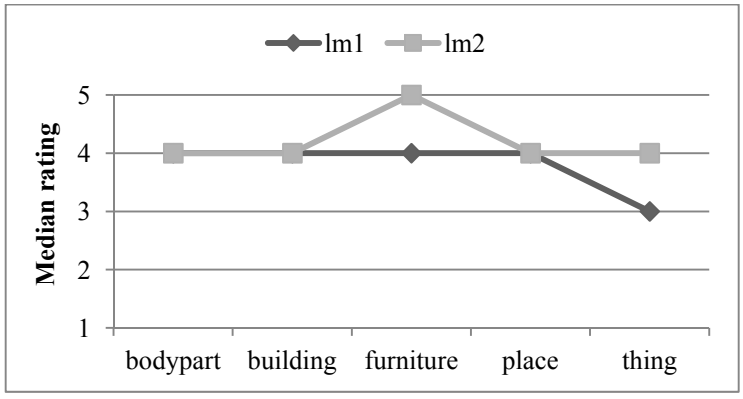


Figure 10. Interaction plot for type of Landmark and word order in picture rating task

This interaction seems to point to the conclusion that the canonical word order sequence where the Trajector phrase precedes the Landmark phrase, i.e. Landmark comes second, is preferred over placing the Landmark phrase first. The reason why the sentences with things as Landmarks where the Landmark phrase precedes the Trajector phrase received low ratings relates to the issue of topic-focus. If we take one of the experimental sentences and pictures that correspond to this experimental condition, e.g. an alarm-clock on the book, we are normally interested in the location of the alarm-clock and hence prefer to place the new information, i.e. the actual location of the alarm-clock, second. It is less natural to start the description with the book and then elaborate that there is an alarm-clock on top of it.

5.1.5 Individual differences between participants

An important issue, which is relevant for reporting experimental results, is the fact that participants differ in their average ratings. As can be seen from Figure 11, there are some participants (4 out of 73) who have a median rating of only 2, but there is also a large number of participants (23 out of 73) who have a median rating of 5.

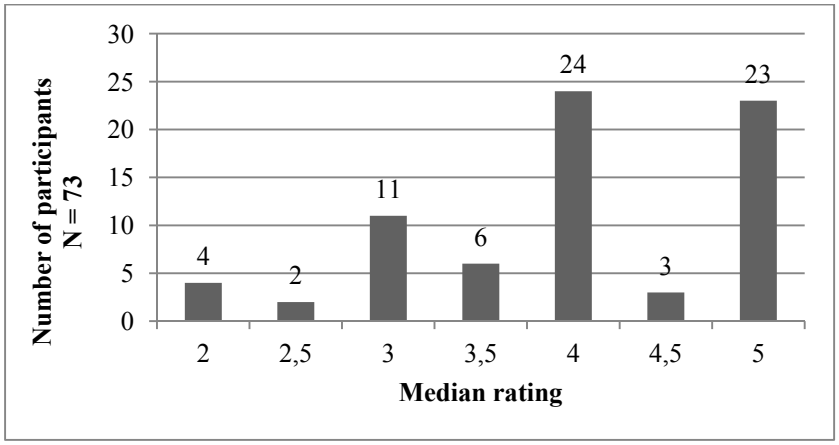


Figure 11. Frequency of participants with different median ratings in picture rating task

In regard to individual differences, it is also interesting to look at the difference between the median ratings averaged over all conditions for the two groups of sentences – adessive sentences and *peal*-sentences – on the level of the individual. These differences are presented in Figure 12, where the x-axis indicates the difference we get if we subtract from the median rating of the adessive sentences the median rating of *peal*-sentences for one and the same individual. Thus, if the difference is -4 as in the first bar of Figure 12, this

means that for these participants the median rating for the adessive sentences was 1 and the median rating for *peal*-sentences was 5, i.e. these participants gave a considerably higher rating for *peal*-sentences when averaged over all observations. The y-axis in Figure 12 indicates how many participants there were in the data set with the difference in median ratings as specified on the x-axis. From Figure 12 it can be seen that there were many participants for whom there was no difference. Averaged over all observations, 27 participants out of the total of 73 had the same median ratings for adessive and *peal*-sentences (the difference is 0 as indicated on x-axis); for 24 participants the difference was only -1, -0.5, 0.5 or 1. However, there were also a small number of individuals for whom the difference is considerable large: for 4 participants the adessive sentences were considerably less acceptable than the *peal*-sentences and for 2 participants the situation is reverse. These results may point to an important conclusion – that there are considerable differences between the use of the adessive construction and the adpositional construction on the level of the individual. This relates to the issue of including Subjects as either fixed or random effects in regression modelling.

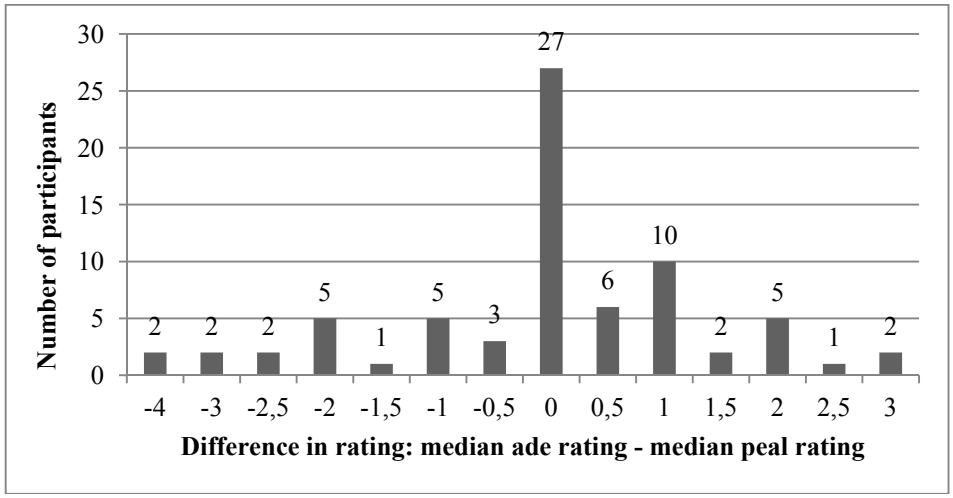


Figure 12. Individual difference in median ratings for adessive and *peal*-sentences in picture rating task

Another relevant issue is the average ratings for the specific 20 landmark-trajector pairs. Although 14 pairs out of 20 received the median rating of 4, there are three pairs ('box on shoulder', 'alarmclock on book', and 'rock on back') that have received a relatively low median rating – 3, and three pairs ('Band-Aid on nose', 'bag on chair', and 'rose on bed') that received a relatively high median rating – 4.5 and 5.

5.1.6 Summary of the exploratory findings

The descriptive findings in the first part of the analysis seem to indicate that both type of Landmark and word order influence the way in which participants rated the sentences. There is an interaction between construction type (i.e. the choice between the adessive and the adposition *peal* 'on') and type of Landmark on the one hand, and an interaction between construction type and word order on the other hand. More specifically, adessive sentences with buildings and places as Landmarks are rated higher than the *peal*-sentences with buildings and places as Landmarks. Conversely, *peal*-sentences with furniture and things as Landmarks receive higher ratings than the adessive sentences with these Landmark types. Both the adessive and *peal*-sentences with bodyparts as Landmarks receive similar ratings. This finding points towards the conclusion that the first five predictions bear out, save for Prediction 2 regarding bodyparts. It was predicted that adessive sentences with bodyparts as Landmarks receive higher ratings than the corresponding *peal*-sentences, but the median ratings are the same for both sentence types. This prediction was made based on the study conducted by Ojutkangas (2008) who looked at the alternation between Finnish interior locative cases and the corresponding adpositions.

As to the interaction between type of construction and word order, there is no difference between the ratings for adessive and *peal*-sentences when the Landmark comes first in the sentences; there is, however, a difference when the Landmark comes second – adessive sentences receive a higher rating than the *peal*-sentences. It was predicted, however, that adessive sentences with Landmark second should receive lower ratings than the corresponding *peal*-sentences. The preliminary results described above are summarised in Table 69 where the values of the two construction types are cross-tabulated with the values of the other explanatory variables. In order to check if any of the differences highlighted might have simply arisen by chance and to test whether the specific predictions are verified or not, the next section presents a more sophisticated statistical analysis of the experimental data.

Table 69. Summary of the exploratory findings of picture rating task

	adessive	**peal 'on'**	**both constructions**
Type of Landmark	building place	furniture thing	bodypart
Word order	lm2		lm1

5.1.7 Repeated measures ANOVA

As we saw from the discussion in section 3.4, there are many multivariate techniques available. The trick, obviously, is in choosing the correct one. Table 70 gives an overview of some of the multivariate techniques that are in essence all applicable to the current analysis – depending, of course, in large part on whether we take the response variable to be continuous or ordinal. In order to appreciate the consequences of using any of the different techniques given in Table 70, the following section provides the analysis of the experimental data using all of these techniques. Although the inclusion of random effects in analysing the data is crucial, the results of using techniques that do not permit the inclusion of random effects are still included in order to demonstrate if and what kind of differences there are in the results arising from this kind of omission. It has to be stressed that the tables comparing the different multivariate methods are not suggesting that when one does not obtain the results one anticipates, one should analyse the data using a different method. The reason behind the comparative tables between different methods is that, in essence, all the techniques employed in the dissertation have been used by various researchers to analyse similar datasets. The aim, therefore, is to see whether these techniques elicit different results and if yes, where the difference lies. At the same time, it has to be stressed that all of the techniques are sufficiently different as to prevent any direct comparisons, i.e. comparisons based on *p*-values or coefficients. However, it is possible to draw inferences about the general trends in the data. First, the description of the results using the more controversial, but more simpler technique – repeated measures ANOVA – is presented, followed by (mixed-effects) logistic regression.

Table 70. Overview of the different multivariate techniques applicable to the analysis of experimental rating data

Multivariate technique	Dependent variable	Random effects included
Repeated measures ANOVA (SPSS)	continuous	yes
Linear mixed-effects regression (R)	continuous	yes
Ordinal logistic regression (SPSS)	ordinal	no
Ordinal mixed-effects logistic regression (R)	ordinal	yes

The repeated measures analysis of variance was carried out using IBM SPSS Statistics 19 software and not R, because the procedure described in Johnson (2008) about how to use R to run a repeated measures ANOVA did not specify how to calculate effect size. The latter is already built into SPSS. Since the data

violated the assumption of sphericity[67], the results of a three-way repeated measures ANOVA are reported with a Greenhouse-Geisser correction. A preliminary analysis of the participants' ratings was conducted using a construction (2) x randomization (4) repeated measures ANOVA. The analysis did not reveal a significant main effect for neither the construction [$F(1, 69) = 0.03$, $p = 0.93$] nor randomization [$F(3, 69) = 0.01$, $p = 0.99$]; nor was the construction x randomization interaction significant [$F(3, 69) = 0.67$, $p = 0.57$]. The data from different versions and randomizations are therefore analysed together in the subsequent analysis.

Following the procedure described in Johnson (2008: 121–134) for a repeated measures analysis of variance, an ANOVA table is constructed with four experimental variables – construction (cx), type of Landmark (lm), word order (wo), and subject (Table 71). In the first column we have the independent variables and their interactions for which we want to know whether they have an effect on the response variable of ratings. The second column lists the degrees of freedom (Df), the third one the sums of squares (Sum Sq), the fourth one gives the mean squares (Mean Sq). The last three columns in Table 71 are of main interest – from there we get the F-values and *p*-values, which we use to decide whether a factor or an interaction has a significant effect on the response variable or not. The last column gives the measure for the effect size – partial eta-squared values[68].

The results indicate that there is a statistically significant although small main effect for type of Landmark [$F(4, 288) = 7.54$, $p < 0.001$, partial $\eta^2 = 0.095$] and word order [$F(1, 72) = 14.99$, $p < 0.001$, partial $\eta^2 = 0.172$]. There is a strong significant effect for the interaction between construction and type of Landmark, $F(4, 288) = 44.26$, $p < 0.001$, partial $\eta^2 = 0.381$; and a relatively strong effect for the interaction between construction type and word order, $F(1, 72) = 21.57$, $p < 0.001$, partial $\eta^2 = 0.231$. The interaction between type of Landmark and word order is significant, but the effect is very small, $F(4, 288) = 3.46$, $p = 0.013$, partial $\eta^2 = 0.046$. There was no main effect of the construction [$F(1, 72) = 0.01$, $p = 0.933$, $\eta^2 < 0.001$], nor was the three-way interaction (construction x type of Landmark x word order) significant [$F(4, 288) = 1.53$, $p = 0.195$, partial $\eta^2 = 0.021$].

[67] The traditional repeated measures ANOVA assumes sphericity. Sphericity relates to the equality of variances of the differences between levels of the repeated measures factor. The violation of sphericity occurs when the variances of the differences between all combinations of the groups are not equal. If the sphericitiy assumption is violated, as in the present case, an approximate test adjusts the degrees of freedom downward for the usual *F* statistic, using the Greenhouse-Geisser adjustment. (Agresti and Finlay 2009: 393).

[68] The values for small, medium, and large EF indexes for analysis of variance are .10, .30, and .50 respectively (Cohen 1992: 157, Table 1).

Table 71. Results for construction (2) x Landmark (5) x word order (2) repeated measures ANOVA (picture rating task)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
construction (cx)	1	0.029	0.0289	0.0071	0.933	0.001
cx:subject	72	291.934	4.0546			
Landmark (lm)	4	25.123	6.2807	7.5406	<0.001	0.095
lm:subject	288	218.303	0.8146			
word order (wo)	1	41.956	41.956	14.995	<0.001	0.172
wo:subject	72	201.456	2.798			
construction:Landmark	4	161.65	40.413	44.262	<0.001	0.381
cx:lm:subject	288	262.95	0.913			
construction:word order	1	18.988	18.9878	21.572	<0.001	0.231
cx:wo:subject	72	63.375	0.8802			
Landmark:word order	4	9.377	2.34418	3.4573	0.009	0.046
lm:wo:subject	288	195.273	0.67803			
construction:lm:wo	4	4.866	1.2164	1.5326	0.1927	0.021
cx:lm:wo:subject	288	228.584	0.7937			

These findings can be explored further by looking at the possible pairwise comparisons – at the moment we only know that there is a significant effect, but we have no idea in which direction it goes. The usual strategy in psycho-linguistics is to perform another repeated measures ANOVA on a subset of the data (Johnson 2008: 128). By taking a subset of the data, one can perform a one-way ANOVA on the subset and still use the subjects as the error term. For example, I looked at the effect of construction in the five different landmark types in five separate repeated measures ANOVAs and found a significant difference between the adessive case and the adposition for all of the five landmark types but one – bodyparts [$F(1,72) = 3.45$, $p = 0.07$, partial $\eta^2 = 0.05$]. There was a small significant difference between the two constructions for building [$F(1,72) = 18.26$, $p < 0.001$, partial $\eta^2 = 0.20$], furniture [$F(1,72) = 12.13$, $p = 0.001$, partial $\eta^2 = 0.14$], and a large significant difference for place [$F(1,72) = 38.47$, $p < 0.001$, partial $\eta^2 = 0.35$], and thing [$F(1,72) = 31.11$, $p < 0.001$, partial $\eta^2 = 0.30$]. These results confirm the exploratory findings of the previous section – adessive sentences with buildings and places receive a higher average rating than the corresponding *peal*-sentences; and *peal*-sentences with pieces of furniture and things receive a higher average rating than the corresponding adessive sentences. The effect for construction type is largest for sentences with places as Landmarks, followed by things, buildings and furniture.

I also looked at the effect of construction with two word order sequences in two further separate repeated measures ANOVAs and found a marginally significant difference between the two constructions in the word order sequence "lm2", i.e. when the Landmark phrase comes second and follows the Trajector phrase [$F(1,72) = 4.30$, $p = 0.04$, partial $\eta^2 = 0.07$], but no significant

construction effects in the word order sequence "lm1", i.e. when the Landmark phrase comes first in the sentence [$F(1,72) = 3.43$, $p = 0.07$, partial $\eta^2 = 0.05$]. This results confirms the exploratory findings above – on average, the adessive sentences received a slightly higher rating than *peal*-sentences when the Landmark followed the Trajector phrase; the two constructions were rated equally high when the Landmark preceded the Trajector.

I also looked at the effect of word order with the five different landmark types in five separate repeated measures ANOVAs and found a significant difference between the "lm1" and "lm2" condition with things [$F(1,72) = 26.99$, $p < 0.001$, partial $\eta^2 = 0.27$], furniture [$F(1,72) = 11.60$, $p = 0.001$, partial $\eta^2 = 0.14$] and a marginally significant effect with body parts [$F(1,72) = 4.79$, $p = 0.03$, partial $\eta^2 = 0.06$] and buildings [$F(1,72) = 4.97$, $p = 0.03$, partial $\eta^2 = 0.06$], but no significant word order effects with places [$F(1,72) = 1.04$, $p = 0.31$, partial $\eta^2 = 0.01$]. This confirms the exploratory findings above – sentences with furniture and things prefer the word order sequence where the Landmark comes second, with the other Landmark types there is no difference – the two word order sequences receive a similarly high rating.

Table 72 summarises the results of repeated measures ANOVA.

Table 72. Comparison of the multivariate techniques used for the analysis of picture rating task (rep. measures ANOVA)

Stat. method	**rep. measures ANOVA**	
	sign.	effect
construction (cx)	not sign.	0.00
landmark (lm)	sign.	0.09
word order (wo)	sign.	0.17
cx*lm	sign.	0.38
cx*wo	sign.	0.23
lm*wo	sign.	0.05
cx*lm*wo	not sign.	0.02
random-effect	yes	
response variable	interval	

There is no significant main effect for construction type – both adessive and *peal*-sentences receive a similarly high rating when generalized over all conditions. The main effects of the other two explanatory variables – type of Landmark and word order – are significant, but the effect size is very small. There is a more pronouncedly significant effect size for the interactions between construction type and landmark, and construction type and word order. There is also a significant effect for the interaction between type of Landmark and word order, but effect is

very small. The three-way interaction between the three explanatory variables is not significant. The analysis will proceed with linear mixed-effect regression, which still treats the response variable as interval, but allows to include a further error term in the model – that of item-based random effects.

5.1.8 Linear mixed-effects regression

The research question we are asking with regression modelling is whether a rating of a sentence in the experiment can be predicted from the locative construction, type of Landmark and the specific word order sequence used in this sentence. Baayen stresses the usefulness of mixed-effects models over the "gold standards" in psycholinguistic studies (Baayen 2008: 259; Baayen, Davidson and Bates 2008; Baayen 2010). Baayen (2008) emphasises that mixed-effects models with crossed random effects are a recent development in statistics. The upside of using this method is that one gets to include subjects in the model as random effects, but the downside is that the dependent variable is still taken to be continuous. In comparing the F-tests (such as the analysis of variance) and mixed-effects modelling, Baayen (2008: 266) reaches the conclusion that mixed-effects models have the same power as the quasi-F test of detecting an effect if it is there; both techniques also possess more or less the same level of risk of incorrectly concluding an explanatory variable is significant. However, Baayen (2008: 266) stresses that mixed-effects models "offer the advantage of being robust with respect to missing data, of allowing covariates to be taken into account, and of providing insight into the full structure of your data, including the random effects".

Following the procedure described in Baayen (2008: 242–259), ratings are modelled as depending on the predictor variables construction, type of the Landmark and word order. In the formula, we also have random-effects terms for Subjects and Items (in this case, the individual sentences used in the questionnaire). Including Subjects and Items as random effects in the model allows fine-tuning with respect to the intercept, i.e. we can lower the intercept for a subject implying that all ratings for that subject become somewhat lower; for subjects who provide high ratings, we may need to increase the intercept, so that all their responses become higher (Baayen 2008: 244). Similarly, some items may be rated lower and we thus may want to adjust the intercept for the individual sentences by means of another random-effects term (Baayen 2008: 245). In R, one can inspect the actual adjustments for specific subjects and specific items with the `ranef()` function (Baayen 2008:246).

First, a model with all the three explanatory variables and all the interactions (including the three-way interaction) was built. Since, however, the three-way interaction between the three explanatory variables and the interaction between type of Landmark and word order turned out to be insignificant, a new model was built excluding these interactions. A new model was built with all the three explanatory variables and the interactions between construction and type of Landmark, and construction and word order.

Figure 13 presents the mixed-effects model object with construction, type of Landmark and word order as fixed-effects and Subjects and Items as random-effects. In the summary of the model object in Figure 13, after a list of summary statistics that describe the quality of the fit of the model to the data, there is a table with the random effects in the model, followed by a table with fixed effects; the summary concludes with a table of correlations of the fixed effects. The part of the summary dealing with fixed effects gives a table of coefficients of the fixed effects – in this case the coefficient for the intercept and for the slope of construction (CX), type of Landmark (LM), word order (WO) and their interactions. The table also lists the associated standard errors, *t*-values, and the *p*-values on the *t*-statistic.

```
Linear mixed model fit by REML
Formula: RATING ~ CX + LM + WO + CX * LM + CX * WO + (1 |
SUBJECT) + (1 |      SENTENCE)
   Data: picture
   AIC   BIC logLik deviance REMLdev
 10199 10289  -5085    10149   10169
Random effects:
 Groups   Name        Variance Std.Dev.
 SENTENCE (Intercept) 0.18821  0.43383
 SUBJECT  (Intercept) 0.19757  0.44449
 Residual             1.74419  1.32068
Number of obs: 2920, groups: SENTENCE, 80; SUBJECT, 73

Fixed effects:
                        Estimate Std. Error t value  Pr(>|t|)
(Intercept)              3.0599     0.1955  15.655    0.0000
CXpeal                   0.4998     0.2664   1.876    0.0607
LMbuilding               0.5874     0.2432   2.415    0.0158
LMfurniture              0.2444     0.2432   1.005    0.3150
LMplace                  0.5875     0.2432   2.416    0.0158
LMthing                 -0.1741     0.2432  -0.716    0.4740
WOLm2                    0.5326     0.1538   3.463    0.0005
CXpeal:LMbuilding       -0.9275     0.3439  -2.697    0.0070
CXpeal:LMfurniture       0.1565     0.3439   0.455    0.6492
CXpeal:LMplace          -1.2680     0.3439  -3.687    0.0002
CXpeal:LMthing           0.4872     0.3439   1.416    0.1567
CXpeal:WOLm2            -0.4226     0.2175  -1.943    0.0521
```

Figure 13. Linear mixed-effects model object (picture rating task)

Before moving on to interpreting the results, the residuals were checked for potential problems with the model specification. It seems that the model is not coping very well with very high and very low ratings. This means that the model does not fit very well and we have to be careful in interpreting the results. Baayen (2008: 258) points out that unlike summaries for simple regression model objects, summary tables for mixed-effects models obtained

with lmer() using linear mixed-effects regression modelling do not list the proportion of variance (R^2) accounted for. This is because there are a number of different sources of variance that are modelled jointly; in addition to the variance explained by fixed effects, we have the variance explained by the by-subject and by-item random effects (Baayen 2008: 258). Baayen (2008: 259) stresses that as is often the case with such experiments, a large proportion of the variance is accounted for just by variability among subjects. He (Baayen 2008: 259) goes on to say that "the method of this kind of data acquisition is inherently very noisy, but the low signal-to-noise ratio is of course exactly the reason why these experiments are generally run with many different subjects and a wide range of items".

The mixed-effects model presented in Figure 13 has three main effects (construction, type of Landmark, and word order) and two interactions (construction x type of Landmark and construction x word order). Since we have interactions in which the main effects are involved, the interpretation of the main effects is not straightforward (Baayen 2008: 166). According to Baayen (2008: 166), in the model with interactions, everything is recalibrated and the main effects by themselves are no longer very informative. This is also why a main effect that is not significant on its own (as construction in the model in Figure 13) is still included in the model as long as it is involved in interactions that are significant.

The experiment reveals a difference between the sentences with different types of Landmarks. Figure 13 confirms what we saw above with the exploratory findings and with the analysis of variance – sentences with buildings and places receive significantly higher ratings than the sentences with bodyparts (the reference level for this variable). The way we can see this from Figure 13 is that the estimates for the predictors construction (LM) with the level *building* (LMbuilding) and *place* (LMplace) given in the second column in the summary output for fixed effects is positive – 0.5874 for the former, and 0.5875 for the latter, indicating scores are predicted to be higher; these result are significant (*t*-values 2.415, 2.416 and *p*-values 0.0158, 0.0158 respectively). The ratings given for furniture and thing do not differ significantly from the ratings given for bodyparts. Word order is also significant predictor according to this model – sequences where the trajector phrase comes second (WOLm2) are rated higher (with a positive estimate of 0.5326) than sentences where the landmark phrase comes first; this result is significant (*t*-value = 3.463, $p < 0.001$). This confirms the results of exploratory findings above, where it was found that the mean rating for "lm2" is 3.73 (SD = 1.44) and the mean rating for "lm1" is 3.39 (SD = 1.55).

The linear mixed-effects regression model presents converging evidence with the exploratory findings and the repeated measures ANOVA for the interactions between construction type and type of Landmark, and construction type and word order. The coefficient for *peal*-sentences (CXpeal) with buildings (LMbuilding) and places as Landmarks (LMplace) have negative coefficients (CXpeal:LMbuilding = -0.9275, CXpeal:LMplace = -1.2680) indicating that

these combinations received significantly (*t*-values -2.697, -3.687 and *p*-values 0.0070 and 0.0002 respectively) low ratings. The coefficient for *peal*-sentences with the word order "lm2" has also a negative coefficient (CXpeal:WOLm2 = -0.4226) indicating that this combination is rated relatively low. However, this finding is only borderline significant: *t*-value = -1.943, *p* = 0.0521.

Following Baayen (2008: 253), the significance of parameters for random effects is assessed by means of a likelihood ratio test, which is carried out by the `anova()` function when applied with two mixed-effects models that have the same fixed-effects, but different number of random-effect parameters. For example, we can evaluate the significance of the by-subject and by-item random effects by fitting a simpler model with only by-subject random intercept that we then compare with the full model. Figure 14 presents the results.

```
anova(pict.lmer1A, pict.lmer1B)
Data: picture
Models:
pict.lmer1B: RATING ~ CX + LM + WO + CX * LM + CX * WO + (1 |
SUBJECT)
pict.lmer1A: RATING ~ CX + LM + WO + CX * LM + CX * WO + (1 |
SUBJECT) + (1 | SENTENCE)
            Df   AIC   BIC  logLik Chisq Chi Df Pr(>Chisq)
pict.lmer1B 14 10305 10389 -5138.6
pict.lmer1A 15 10179 10269 -5074.5 128.2      1  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 14. ANOVA of the linear mixed-effects models with and without sentences as random effects (picture rating task)

From Figure 14 it can be seen that the associated probability of the chi-squared test is small hence the additional parameter (random-effect parameter for by-item random effect) in the more complex model is justified: the model with by-item random effect fits the data significantly better.

The results of the repeated measures ANOVA and linear mixed-effects regression model are presented in Table 73. Table 73 demonstrates that the results produced by the two techniques are, in large part converging. The results diverge on one interaction – between the type of landmark and word order. While the results of a repeated measures ANOVA indicate that the interaction is significant (although the effect size is small), the results of linear mixed-effects regression modelling lead us to conclude that this interaction is not significant. This may be due to several reasons. One possible explanation is that the model produced with linear mixed-effects regression includes both subjects and items as random effects, while the analysis of variance only takes into account the repeated measures by subjects. At the same time, we have to bear in mind that the linear mixed-effects regression model has some problems with the model fit which renders the results less reliable. I will now move on to analysing the data

by employing ordinal logistic regression. First the results of a simple ordinal logistic regression (using SPSS) are presented, followed by a very recent technique –ordinal mixed-effects logistic regression.

Table 73. Comparison of the multivariate techniques used for the analysis of picture rating task (rep. measures ANOVA vs. linear mixed-effects reg.)

Stat. method	rep. measures ANOVA		Linear mixed-effects regression
	sign.	effect	
construction (cx)	not sign.	0.00	not sign.
landmark (lm)	sign.	0.09	sign.
word order (wo)	sign.	0.17	sign.
cx*lm	sign.	0.38	sign.
cx*wo	sign.	0.23	sign.
lm*wo	sign.	0.05	not sign.
cx*lm*wo	not sign.	0.02	not sign.
random-effect	yes		yes
response variable	interval		interval

5.1.9 Ordinal logistic regression

While the two previous statistical techniques described above (repeated measures ANOVA and linear mixed-effects modelling) assume that the response variable is continuous, the remaining two multivariate techniques described here take the response variable to be ordinal rather than continuous. For the purposes of comparison, I will first present the data analysis where the issue of random-effects for subjects is ignored and then use a new method that allows the inclusion of random-effects also in ordinal logistic regression modelling[69]. The first follows the procedure described in Norušis (2011: 69–89) and the analysis itself will be carried out in SPSS; the second follows the procedure described in Divjak (in progress) and the analysis is carried out in R.

With ordinal logistic regression the research question is whether a rating given to a sentence can be predicted from the combination of variables: construction type, type of Landmark and word order. As we saw from section 3.4.4, ordinal regression models are specialised cases of the general linear model. The coefficients in the logistic regression model tell you how much the logit changes based on the values of the predictor variables (Norušis 2011: 70). Instead of considering the probability of an individual event, you consider the probability of that event and all events that are ordered before it (Norušis 2011: 70).

[69] Currently, only a single random term is allowed in the model.

The first step in model building is fitting the full model with rating as the response variable and construction type, type of Landmark, word order and their interactions as factors or predictor variables. First, a full model with all the variables and their interactions was fitted. Since the interaction between type of Landmark and word order was not significant, it was left out of the model. The three-way interaction between the variables, however, was significant, so it was left in the model. Before interpreting the results, the assumption of the parallel lines or the parallel odds assumption, and other model criticism needs to be checked. The model assumption is that the regression coefficients are the same for all five categories. Since the observed significance level in Table 74 is large, there is not sufficient evidence to reject the parallelism hypothesis, i.e. the model does not violate the assumption.

Table 74. Test of parallel lines for ordinal logistic regression model (picture rating task)

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	424.340			
General	368.400	55.940	57	.515

In assessing the model fit for ordinal logistic regression models in SPSS, various statistics are available (see section 3.4.1 above). One way of assessing whether the model fits well, is by looking at the difference between observed cell counts and expected cell counts. Table 1 in Appendix 3 presents the observed and expected (given in brackets) cell counts for all the experimental conditions by the 5 rating categories. There does not seem to be much difference between the observed and expected counts. The Pearson and Deviance goodness-of-fit statistics given in Table 75 confirm this – the observed and expected cell counts are similar and we can conclude that the model fits well.

Table 75. Goodness-of-fit statistics for the ordinal regression model (picture rating task)

	Chi-Square	df	Sig.
Pearson	54.245	57	.579
Deviance	55.940	57	.515

Continuing with the model criticism, one should also pay attention to the strength of association. According to Norušis (2011: 81) there are several R^2-like statistics that can be used to measure the strength of the association between the response variable and the predictor variables. But Norušis (2011: 81) also points out that they "are not as useful as the R^2 statistic in regression, since their interpretation is not straightforward". For the experimental data at hand, the values of the pseudo *R*-square statistics are small as can be seen from Table 76.

Table 76. Pseudo R-square statistics for the ordinal logistic regression model (picture rating task)

Cox and Snell	.076
Nagelkerke	.080
McFadden	.027

Considering that the model diagnostic figures are within the acceptable range, it can be assumed that the predictions made by the model can be trusted and we can proceed with looking at the actual results. The output of the analysis done in SPSS is presented in Table 77. The output gives the estimated coefficients for the model. As is the case with categorical predictors in models with intercepts, the number of coefficients displayed is one less than the number of categories of the variable and in SPSS the reference category is the last category (cf. R where the reference category is the first category). Larger coefficients indicate an association with larger scores; alternatively, negative coefficients indicate that lower scores are more likely. The Wald statistic is the square of the ratio of the coefficient to its standard error. Small significance level indicates that the null hypothesis that the ratio is zero can be rejected.

Table 77. Output for the ordinal logistic regression model (picture rating task)

	Esti-mate	Std. Error	Wald	Df	Sig.
[RATING=1]	-2.408	.165	213.91	1	.000
[RATING=2]	-1.467	.160	83.66	1	.000
[RATING=3]	-.897	.159	31.88	1	.000
[RATING=4]	-.198	.158	1.56	1	.211
[CX=ade]	-.451	.218	4.27	1	.039
[LM=bodypart]	-.599	.217	7.59	1	.006
[LM=building]	-.843	.216	15.19	1	.000
[LM=furniture]	.175	.226	.60	1	.437
[LM=place]	-1.278	.216	35.01	1	.000
[WO=lm1]	-.217	.220	.97	1	.324
[CX=ade]*[LM=bodypart]	.600	.304	3.89	1	.049
[CX=ade]*[LM=building]	1.403	.309	20.66	1	.000
[CX=ade]*[LM=furniture]	.075	.312	.06	1	.809
[CX=ade]*[LM=place]	1.811	.308	34.52	1	.000
[CX=ade]*[WO=lm1]	-1.001	.305	10.76	1	.001
[CX=ade]*[LM=bodypart]*[WO=lm1]	.433	.298	2.11	1	.146
[CX=ade]*[LM=building]*[WO=lm1]	.779	.306	6.49	1	.011
[CX=ade]*[LM=furniture]*[WO=lm1]	.536	.300	3.18	1	.074
[CX=ade]*[LM=place]*[WO=lm1]	.864	.306	7.98	1	.005
[CX=peal]*[LM=bodypart]*[WO=lm1]	.365	.306	1.42	1	.233
[CX=peal]*[LM=building]*[WO=lm1]	.111	.303	.13	1	.714
[CX=peal]*[LM=furniture]*[WO=lm1]	-.150	.314	.23	1	.633
[CX=peal]*[LM=place]*[WO=lm1]	.194	.303	.41	1	.521

There appears to be a relationship between construction type, type of Landmark and the ratings. As was noted above, since there are significant interactions in which the main effects are involved, the interpretation of the main effects is not straightforward. From the perspective of the research question, studying the main effects is not informative either, since the main interest lies in the interactions between the type of construction and the other two variables – type of Landmark and word order. The results show that there is a relationship between the interaction of construction with type of Landmark and rating. For any rating level, adessive sentences with body parts, buildings and places are rated higher than adessive sentences with things and *peal*-sentences with body parts, buildings, furniture, place, and thing. Ratings and the interaction of construction type with word order are also related – adessive sentences, where the Landmark phrase comes first are rated lower than adessive sentences, where Landmark comes second or peal-sentences with both "lm1" and "lm2".

The results obtained with ordinal logistic regression can be compared to the repeated measures ANOVA and linear mixed-effects regression modelling. The results of the three techniques are presented in Table 78. Again there are both similarities and differences. All three techniques indicate that there is an effect of type of landmark (bearing in mind that with regression models where significant interactions are involved, the interpretation of the main effects is not straightforward). The ordinal regression model also shows a significant effect of construction type, but not of word order unlike the other two techniques. All three techniques confirm that there is a significant interaction between the type of construction and type of landmark and between the type of construction and word order. The ordinal logistic regression model confirms the results of the linear mixed-effects regression model in that the interaction between type of Landmark and word order is not significant. Differently from the other two techniques, ordinal logistic regression model showed a significant effect for the three-way interaction.

Table 78. Comparison of the multivariate techniques used for the analysis of picture rating task (rep. measures ANOVA vs. linear mixed-effects reg. vs. ordinal log. reg.)

Stat. method	rep. measures ANOVA		Linear mixed-effects regression	Ordinal logistic regression
	sign.	effect		
construction (cx)	not sign.	0.00	not sign.	sign.
landmark (lm)	sign.	0.09	sign.	sign.
word order (wo)	sign.	0.17	sign.	not sign.
cx*lm	sign.	0.38	sign.	sign.
cx*wo	sign.	0.23	sign.	sign.
lm*wo	sign.	0.05	not sign.	not sign.
cx*lm*wo	not sign.	0.02	not sign.	sign.
random-effect	yes		yes	no
response variable	interval		interval	ordinal

5.1.10 Ordinal mixed-effects logistic regression

This specific method, although highly desirable for the kinds of experimental data discussed in the present thesis, is fairly new and the R package `ordinal` for conducting this analysis came out only in 2010[70]. The two main advantages of this method are that it, first and foremost, treats the ratings as ordinal and secondly, it allows the inclusion of random-effects. The disadvantage of this method is that since it is fairly new, there are no textbooks available that explain the procedure. The only available document is the R manual (available at: http://cran.r-project.org/web/packages/ordinal/ordinal.pdf). The only linguistic study that uses this method is to my knowledge Divjak (in progress). As the technique is still new, there are as yet not many model criticism measures available. The measures reported here are based on Divjak (in progress).

The model was built stepwise, starting with the most complex model that included all the predictors (construction type, type of Landmark, and word order) and their interactions. Since the two-way interaction between type of Landmark and word order and the three-way interaction is not significant, these were taken out of the model. The best model, presented in Figure 15, contains the random effect for Subject and as fixed effects the three predictors and the interaction between construction type and type of Landmark and the interaction between construction type and word order. The results indicate that all three predictor variables (construction type, type of Landmark, and word order) are significant. The interaction between construction type and type of Landmark, as well as the interaction between construction type and word order are also significant.

One diagnostic[71] that the `ordinal` package in R currently does provide is the condition number of the Hessian measuring the empirical indentifiability of the model (Divjak in progress). The higher the number, the more ill-defined the model, but there is no agreed upon cut-off level for good vs. poor fit. Christensen (2011: 3) points out that high numbers, e.g. larger than 10^4 or 10^6 indicate that the model does not fit well. The condition number of Hessian for this model equals 617.2845, i.e. is lower than 10^4, and signals that the present model predicts the ratings relatively well using the combination of explanatory variables included in the model.

[70] For example, the current version of SPSS does not yet allow to build ordinal regression models with random variables. This demonstrates the usefulness of using R in doing the analysis of the results. While SPSS is more userfriendly, R has the benefit of being more flexible in incorporating new techniques.

[71] There are other diagnostics available, but since they are not implemented in the package itself, they need to be calculated manually.

```
Cumulative Link Mixed Model fitted with the Laplace approximation

Call:
clmm(location = as.factor(RATING) ~ LOCPH * TYPE + LOCPH * WO,
    random = as.factor(SUBJECT), data = pict.all, Hess = TRUE,
    link = "logistic")

Random effects:
                        Var  Std.Dev
as.factor(SUBJECT) 0.4437718 0.666162

Location coefficients:
                      Estimate Std. Error z value Pr(>|z|)
CXpeal                 0.6507   0.1684     3.8647  .000
LMbuilding             0.7996   0.1579     5.0644  .000
LMfurniture            0.3336   0.1542     2.1643  .030
LMplace                0.8199   0.1579     5.1912  .000
LMthing               -0.2188   0.1533    -1.4278  .153
WOLm2                  0.7473   0.0995     7.5122  .000
CXpeal:LMbuilding     -1.1628   0.2196    -5.2960  .000
CXpeal:LMfurniture     0.2524   0.2183     1.1561  .248
CXpeal:LMplace        -1.6499   0.2197    -7.5109  .000
CXpeal:LMthing         0.6911   0.2174     3.1790  .002
CXpeal:WOLm2          -0.6451   0.1389    -4.6430  .000


No scale coefficients

Threshold coefficients:
    Estimate Std. Error z value
1|2 -1.3082   0.1471    -8.8952
2|3 -0.3206   0.1443    -2.2226
3|4  0.2889   0.1442     2.0033
4|5  1.0538   0.1453     7.2534

log-likelihood: -4116.422
AIC: 8264.844
Condition number of Hessian: 617.2845
```

Figure 15. Mixed-effects ordinal logistic regression model (picture rating task)

5.2 Discussion of the results: picture rating task

An overview of the results of all four multivariate techniques used to analyse the results of the picture rating task are given in Table 79. The results that are the same across all techniques are in bold.

Table 79. Comparison of the multivariate techniques used for the analysis of picture rating task

Stat. method	rep. measures ANOVA		linear mixed-effects regression	ordinal logistic regression	ordinal mixed-effects log. regression
	sign.	effect			
construction (cx)	not sign.	0.00	not sign.	sign.	sign.
landmark (lm)	**sign.**	**0.09**	**sign.**	**sign.**	**sign.**
word order (wo)	sign.	0.17	sign.	not sign.	sign.
cx*lm	**sign.**	**0.38**	**sign.**	**sign.**	**sign.**
cx*wo	**sign.**	**0.23**	**sign.**	**sign.**	**sign.**
lm*wo	sign.	0.05	not sign.	not sign.	not sign.
cx*lm*wo	not sign.	0.02	not sign.	sign.	not sign.
random-effect	yes		yes	no	yes
response variable	interval		interval	ordinal	ordinal

All four techniques indicate that there is a main effect of one out of three explanatory variables – type of Landmark is significant, while construction type and word order are not. In addition, all models also indicate that the two-way interactions between construction type and type of Landmark, and construction type and word order are significant. There is, however, diverging evidence for the two-way interaction between type of Landmark and word order, and for the three-way interaction. The two-way interaction is significant according to the repeated measures ANOVA (the effect size is very small, however), while the other three techniques do not confirm this. The three-way interaction is significant according to the simple ordinal logistic regression analysis, but the other multivariate techniques do not confirm this. The simple ordinal logistic regression model is, in fact, the one that diverges the most from the other models. This may be due to the fact that it does not include either subject or items as random effects. Since the diagnostics for the linear mixed-effects model showed that the model does not fit the data very well, the ordinal mixed-effects logistic regression model and repeated measures ANOVA are taken to be the most trustworthy. It should be born in mind, however, that with the repeated measures ANOVA the response variable is (somewhat erroneously) treated as interval rather than ordinal.

It was predicted that the adessive sentences with bodyparts, buildings, and places receive higher ratings than the corresponding *peal*-sentences; and that the *peal*-sentences with things and furniture receive higher ratings than the corresponding adessive sentences. All four multivariate techniques confirm that the interaction between construction type and type of Landmark is significant. The repeated measures ANOVA indicates that the effect is relatively large (partial $\eta^2 = 0.38$). The exploratory results and the separate repeated measures ANOVAs confirm that save for bodyparts, the other four predictions are confirmed. The ratings for adessive and *peal*-sentences with bodyparts did not differ from each other significantly. Generalising these results to the findings reported in the literature (e.g. Bartens 1978, Ojutkangas 2008), we can say that larger, static locations (like buildings and places) prefer the synthetic case construction and smaller, manipulable objects (like pieces of furniture and things) prefer the analytic adpositional construction. These results also confirm the results of the two corpus studies described above where a similar conclusion is drawn.

It was furthermore predicted that the adessive sentences where the Landmark phrase precedes the Trajector phrase receive higher ratings than the corresponding *peal*-sentences and that the *peal*-sentences where the Landmark phrase follows the Trajector phrase receive higher ratings than the corresponding adessive sentences. All four multivariate techniques confirm that the interaction between construction type and word order is significant, although as the repeated measures ANOVA demonstrates, this effect is smaller (partial $\eta^2 = 0.23$) than the effect for the interaction between construction type and type of Landmark. The exploratory findings and the follow-up repeated measures ANOVAs demonstrate that there is no significant difference between the ratings for adessive and *peal*-sentences when the Landmark comes first; there is, however, a significant difference when the Landmark comes second – adessive sentences receive higher ratings than the corresponding *peal*-sentences.

These findings indicate that the predictions pertaining to word order (i.e. that the adessive construction prefers the "lm1" sequence and the *peal*-construction the "lm2" word order) are not borne out. The reason for the opposite result may be due to the ambiguity between the use of the adessive case in the locative construction and the possessive construction. As we saw in Chapter 2, the Estonian possessive construction is comprised of the possessor inflected for the adessive case and this element of the construction normally comes at the beginning of the construction. In order to disambiguate between the two readings, it is preferred by language users if the Landmark phrase expressing location comes after the Trajector phrase, and not before it as in the possessive construction. Another possible explanation is of a more general nature and pertains to information structure. It is simply preferred that the Landmark phrase should follow the Trajector phrase, since we are interested in the location of the Trajector and new information comes at the end of the clause in Estonian, like in many other languages. For *peal*-sentences, however, there is no difference – whether the Landmark phrase comes first or second is not important, both conditions receive an equally high average rating.

5.3 Results of the sentence rating task

Before presenting the results of the sentence rating task, a few key aspects about the design are in order as a reminder; the full details are given above, in section 3.2.2. The sentence rating task is a 2 (type of construction: adessive vs. *peal* 'on') x 2 (length of the Landmark phrase: short vs. long) x 2 (relative word order between Trajector and Landmark: Landmark first vs. Landmark second) factorial design. For each of the 8 experimental conditions, 4 different lexicalizations of the Landmark and Trajector were used. In addition, 8 control sentences were included to control for the effect of different lexicalizations of the Landmark and Trajector phrases. The response variable is a rating given on a scale from 1 ('not at all probable') to 7 ('very probable') by the participants who were asked to rate how probable it is that they themselves would use the sentence presented in the questionnaire. The questionnaire was distributed electronically using the eFormular tool and 103 subjects participated in this task. Every participant rated all of the 40 (32 experimental + 8 control) sentences.

The sentence rating task tested the influence of two morphosyntactic variables: length of the Landmark phrase and word order. Both of these variables were included in the multivariate corpus study and the results show that while the length of the Landmark phrase is one of the strongest variables in predicting the choice between the adessive and *peal*-construction, the relative word order between the Landmark and Trajector is only significant when it is factored in with other variables (length, complexity, and mobility of Landmark, word class of Trajector, and verb group). In order to verify that there is a significant association between length and construction type, and that language users are sensitive to it, and to check if and how word order influences the choice, the sentence rating task was designed. The results of this task are presented below. The presentation of the results follows the same logic as above for the results of the picture rating task. First an exploratory overview of the essential nature of the dataset is given, followed by more sophisticated statistical analyses.

5.3.1 Description of the results

A similar trend of the data being skewed toward higher ratings, as seen in the picture rating task, can also be seen for the data set of the sentence rating task. Figure 16 presents the frequency of each rating category on a scale from 1 to 7 as attested in the data (averaged over all observations). Compared to the data from the picture rating task, where the highest rating category comprised 42% of the total ratings, the cut-off in this dataset is less sudden – the highest rating category, 7, is still the biggest (i.e. the highest bar in Figure 16), but it takes up only 24% of the total of 3,296 observations.

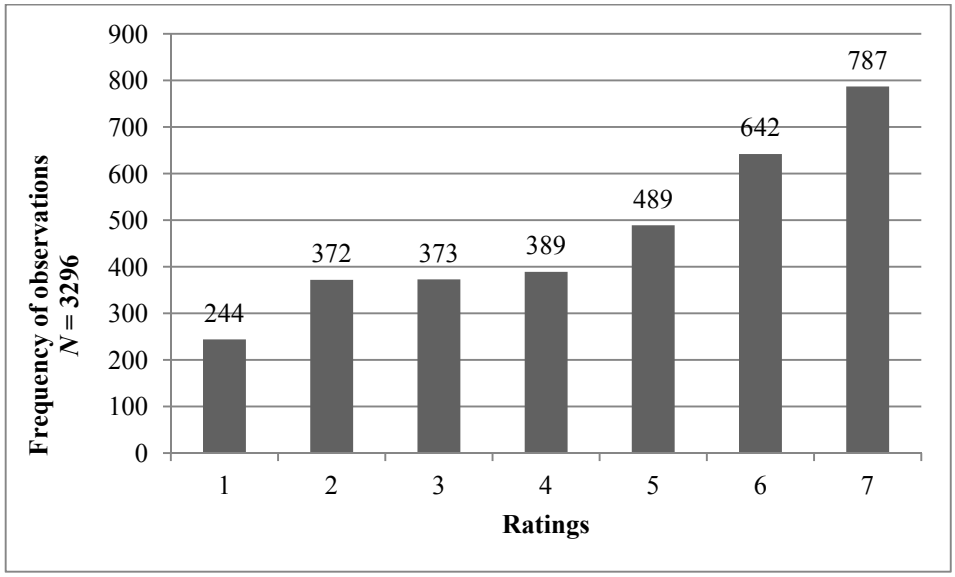


Figure 16. Frequency of ratings averaged over all observations in sentence rating task

As pointed out in discussing the results of the picture rating task, the measures of central tendency for ordinal data are the median (the value separating the higher half of the data sample from the lower half) and the mode (the value that occurs most frequently in a data set). The central tendency for interval data is the mean, but again we should remember that for a skewed data set, the median is a more appropriate measure. In order to show what kind of differences looking at different measures of central tendency entails, Table 80 provides all of these measures for all of the 8 conditions. Table 80 is arranged in alphabetical order. Since we see from Figure 16 that the most frequently given rating is 7, it is not surprising that for five out of the eight conditions the mode is 7. It is, however, noteworthy that all of the three conditions for which the mode was lower than seven, are with the adessive case construction. All of the sentence types with the *peal*-construction are rated higher or receive a similarly high rating as the corresponding sentence types with the adessive construction. The lowest rating was assigned to adessive sentences with long landmark phrases at the beginning of the sentence ("ade_long_lm1").

Table 80. Median acceptability ratings for all experimental conditions in sentence rating task

Condition	Median	Mode	Mean	SD
ade_long_lm1	**3**	2	3.57	1.94
ade_long_lm2	**4**	5	3.95	1.88
ade_short_lm1	6	7	5.33	1.85
ade_short_lm2	5	6	4.50	1.87
peal_long_lm1	5	7	4.46	1.91
peal_long_lm2	5	7	4.82	1.89
peal_short_lm1	6	7	5.83	1.47
peal_short_lm2	6	7	5.09	1.82

For clarity and a better overview of the data, information given in Table 80 is presented visually in Figure 17, where the conditions are arranged not in alphabetical order as in Table 80, but from lowest to highest rating. Figure 17 was produced using the median as the measure for the average tendency.

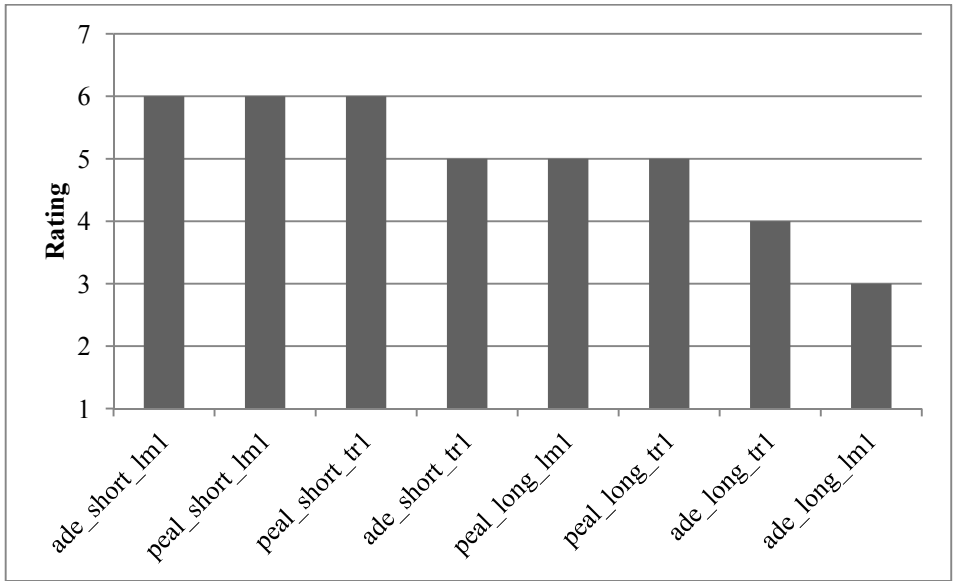


Figure 17. Median acceptability ratings for all experimental conditions in sentence rating task

One of the crucial aspects that Table 80 and Figure 17 suggest is that, differently from the results obtained in the picture rating task, the data points for the

two constructions seem to be differently distributed in the sentence rating task – *peal*-sentences are rated higher than the adessive sentences. This intuition is further confirmed if we look at Figure 18 which presents the frequency counts for the two sentence groups according to the seven rating categories.

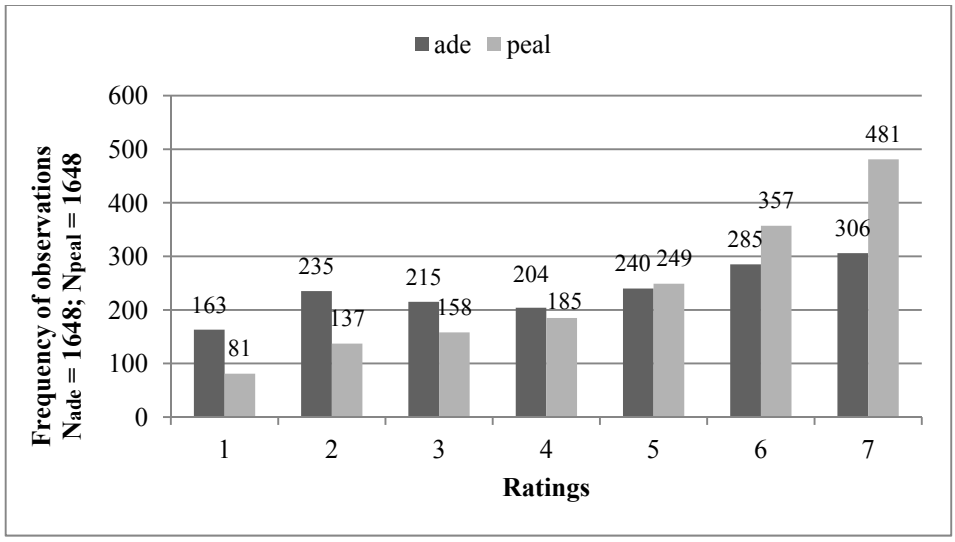


Figure 18. Frequency of ratings for adessive and *peal*-sentences averaged over all observations in sentence rating task

Figure 18 shows that the distribution of the data points is not the same for the two groups of observations. While the bars for the *peal*-construction are higher on the right hand side of the graph where the higher rating categories are, the left-hand side of the graph is different – the bars for the adessive construction are higher instead. While the distribution of the adessive sentences has a more symmetric distribution, the distribution of *peal*-sentences is heavily skewed – as seen from Figure 18, there is a large number of *peal*-sentences that received a rating of 7 (481), while the number of corresponding adessive sentences was 306. For the lowest rating category, 163 adessive observations fell into this category, while the frequency for the *peal*-construction in this category was only 81. These figures point to the conclusion that, averaged over all conditions, there is a difference between the adessive sentences and *peal*-sentences.

When averaged over all observations, the mode rating for the adessive sentences and *peal*-sentences did not differ – it was 7 in both cases. Nevertheless, the median and mean ratings did differ. The median rating was 5 for the adessive sentences and 6 for the *peal*-sentences, suggesting that the adessive case is a less likely candidate out of the two constructions to be used with the types of sentences included in this rating task. The mean ratings show the same

result – the mean rating for the *peal*-sentences (5.05, SD = 1.85) is higher than for the corresponding adessive sentences (4.34, SD = 2.00). This difference is statistically significant – both a Wilcoxon (V = 707.5, $p < .001$) and a *t*-test for dependent samples ($t = -5.81$, $df = 102$, $p < .001$) confirm this.

Exploratory analysis of the data (mainly graphical examinations) precedes the sections on inferential statistics (5.3.7 – 5.3.10). However, let it be said in anticipation that the different inferential statistical methods employed confirm that all of the main effects (construciton type, length of the Landmark phrase, and word order) are significant, as is the interaction between length and word order.

5.3.2 Construction type x length of the Landmark phrase

Before we look at the interaction itself, it may be interesting to look at the central tendencies for sentences with short and long Landmark phrases. It is crucial to bear in mind that in both conditions, the sentences were of equal length in terms of total number of words – each sentence was five words long. If the Landmark phrase was short, i.e. a bare phrase without modifiers, then the length of the whole sentence was kept at five by making the Trajector phrase long. On the other hand, when the Landmark phrase was long (modified by two adjectives), the Trajector phrase was kept short and the whole sentence was again five words long. The mode for sentences with long Landmark phrases was 6 and for short Landmark phrases 7; the corresponding medians were 4 and 6 and the corresponding means 4.20 (SD = 1.96) and 5.19 (SD = 1.82). So even though the total length of the sentence was kept constant, people seem to prefer short Landmark phrases with long Trajector phrases over long Landmark phrases with short Trajector phrases.

We will now move on to the interaction between the locative construction and the length of the Landmark phrase. Generalised over all observations, Table 81 provides the medians, modes, means and standard deviations for the long and short Landmark phrases with the two constructions.

Table 81. Median, mode and mean ratings for the long and short Landmark phrases with the two locative constructions in sentence rating task

Length of Lm phrase	adessive sentences				peal-sentences			
	Median	Mode	Mean	SD	Median	Mode	Mean	SD
long	**4**	2	3.76	1.92	**5**	7	4.64	1.91
short	**5**	7	4.91	1.91	**6**	7	5.46	1.67

The results do not indicate that there is an interaction involved – for both constructions, shorter Landmark phrases are preferred over longer Landmark phrases. Instead, the central tendencies reported in Table 81 once again indicate what was seen above – adessive sentences received lower ratings than *peal*-sentences. As seen from Figure 19, there is no interaction. Instead, the effects of the two variables – type of construction and length of the Landmark phrase – are additive. That means that the difference in the median ratings between sentences with the adessive construction and those with the *peal* construction remains –1 regardless of whether the landmark phrase is long (4 – 5 = –1) or short (5 – 6 = –1).

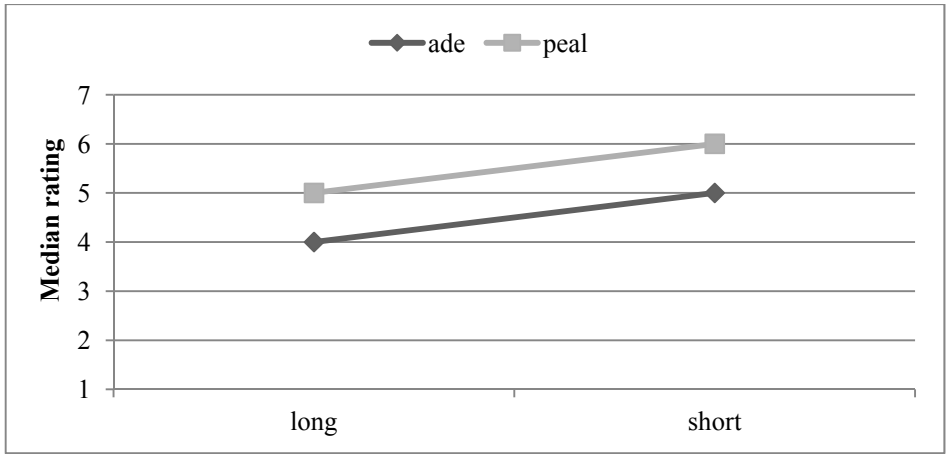


Figure 19. Interaction plot for type of construction (adessive vs. *peal*) and length of the Landmark phrase in sentence rating task

It was predicted that the adessive case construction will be preferred with longer Landmark phrases and *peal*-construction with shorter Landmark phrases. The prediction about *peal*-sentences being preferred with short Landmark phrases seems to hold. However, the combination of adessive construction with a long Landmark phrase was rated the lowest – the opposite of what was predicted. I will come back to discussing the reasons why this may be so in the section on general discussion, but it may be said at this point that the results in all probability reflect the limitations of the stimuli. The experimental stimuli only included one type of Landmarks. All Landmarks denoted pieces of furniture, i.e. relatively small, manipulable objects: 'a chair', 'a table', 'a couch', and 'a bed'. As described in the experimental design in section 3.2.2, one of the aims of this experiment was to keep the semantic variable 'type of Landmark' constant and the selection of these specific Landmarks was intentional. However, as we saw from the results of the picture rating study, the type of Landmark was a significant semantic predictor and pieces of furniture tend to prefer the ad-positional construction over the case construction.

5.3.3 Construction type x word order

Generalised over all observations, the median ratings for the two word order sequences – sentences where the landmark comes first ("lm1") vs. where it comes second ("lm2") – was the same for both (5). However, both the mode and the mean rating was higher for the "lm1" sequence (mode = 7, mean = 4.80, SD = 2.00) compared to the mode and the mean rating for "lm2" (mode = 6, mean = 4.59, SD = 1.91). It was predicted that there is an interaction between the two locative constructions and the two word order sequences: adessive is preferred with "lm1" and *peal* with "lm2". Table 82 gives the medians, modes, means and standard deviations for the two word order sequences according to the two locative constructions (generalised over all observations). Again, we see a similar result as with the length of the Landmark phrase – adessive sentences are rated lower than *peal*-sentences.

Table 82. Median, mode and mean ratings for the two word order sequences according to the two locative constructions in sentence rating task

Word order	adessive sentences				peal-sentences			
	Median	Mode	Mean	SD	Median	Mode	Mean	SD
lm1	**5**	7	4.45	2.09	**6**	7	5.15	1.84
lm2	**4**	6	4.22	1.89	**5**	7	4.96	1.86

Figure 20 illustrates another case where instead of an interaction, there is an additive effect. Although there are differences in the medians, modes and means (cf. Table 82) between these two constructions, participants gave the adessive sentences with both word order sequences lower ratings than the corresponding *peal*-sentences. Discussion as to why this may be so is deferred and will be taken up in the summary paragraph at the end of this section.

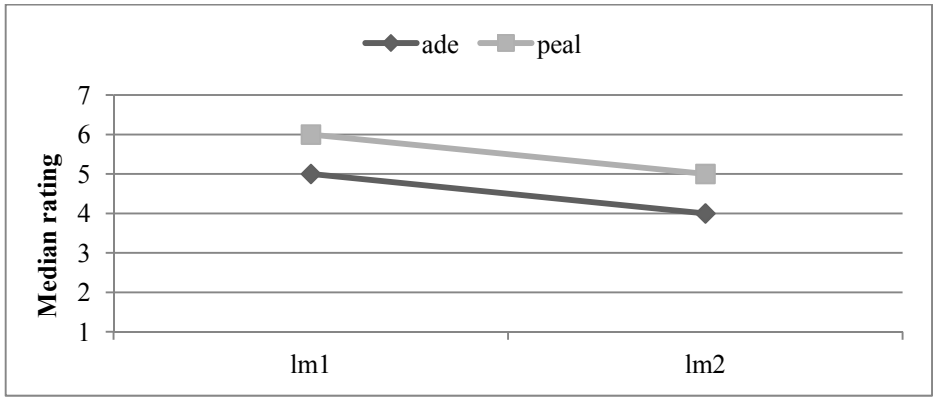


Figure 20. Interaction plot for type of construction (adessive vs. *peal*) and word order in sentence rating task

5.3.4 Length of Landmark phrase x word order

In addition to the previous two-way interactions for which I had specific predictions, the experimental design allows taking a look at a third two-way interaction – that of word order and the length of the Landmark phrase. The medians, modes, means and standard deviations are presented in Table 83.

Table 83. Median, mode and mean ratings for long and short Landmark phrases according to the two word order sequences in sentence rating task

Length of Lm phrase	lm1				lm2			
	Median	Mode	Mean	SD	Median	Mode	Mean	SD
long	**4**	3	4.01	1.98	**5**	6	4.38	1.93
short	**6**	7	5.58	1.69	**5**	6;7	4.79	1.86

Unlike the previous two graphs, Figure 21 presents an interaction, this time between length of the Landmark phrase and word order. The median ratings for sentences where the Landmark phrase comes second are the same for both long and short Landmark phrases. However, there is a sharp contrast between the medians, modes and means between the sentences where the Landmark comes first and the Landmark phrase is long and between the sentences where it is short. Sentences with long Landmark phrases with the Landmark first receive considerably lower ratings (median 4, mean 4.01, SD = 1.98) than sentences with short Landmark phrases and the Landmark first (median 6, mean 5.58, SD = 1.69). The reader is reminded that all sentences contained an equal number of words – the shortness of the Landmark phrase is compensated by making the Trajector phrase longer. These results can be explained by information structure – longer Landmark phrases, where one would expect to find new information, are preferred to come after the Trajector phrase.

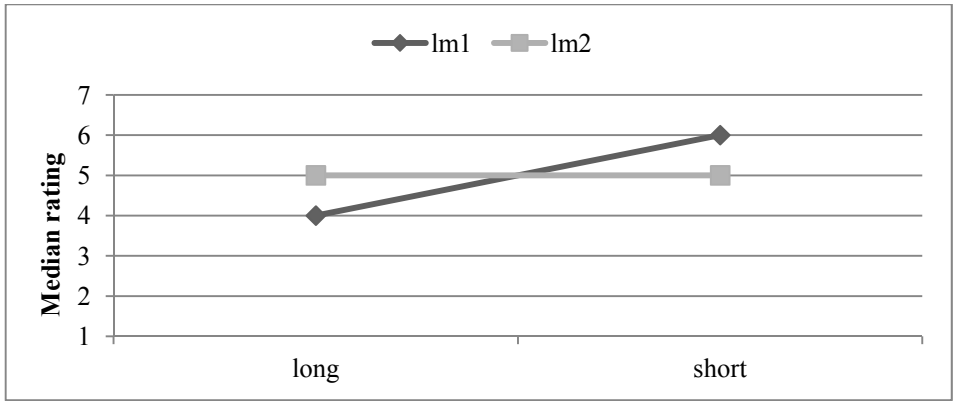


Figure 21. Interaction plot for the length of Landmark phrase and word order in sentence rating task

5.3.5 Individual differences

Individuals differ in their average ratings. Figure 22 presents the frequency of subjects with the same median ratings – the x-axis corresponds to the seven rating categories and the bars show how many individuals had this specific median rating when generalised over all observations. As can be seen from Figure 22, there are some participants (5 out of 103) who have a median rating of only 2; there is also an equal number of participants (5 out of 103) who have a median rating of 7.

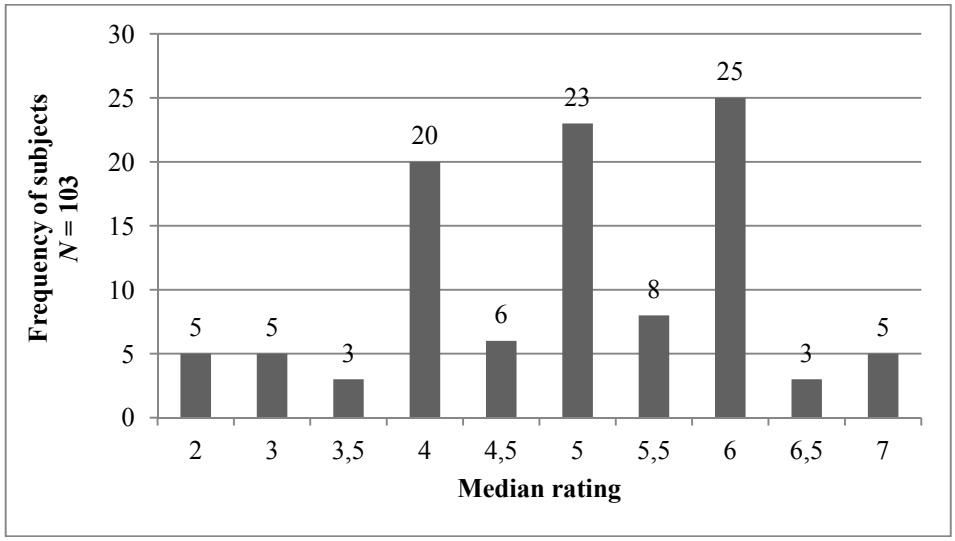


Figure 22. Frequency of participants with different median ratings in sentence rating task

In regard to individual differences, it is also interesting to look at the difference between the median ratings averaged over all conditions for the two groups of sentences – adessive sentences and *peal*-sentences – on the level of the individual. These differences are presented in Figure 23, where x-axis indicates the difference we get if we subtract from the median rating of the adessive sentences the median rating of *peal*-sentences for one and the same individual. Thus, if the difference is -6 as in the left end of the bar plot in Figure 23, this means that for these participants the median rating for the adessive sentences was 1 and the median rating for *peal*-sentences was 7, i.e. these participants gave a considerably higher rating for *peal*-sentences when averaged over all of the conditions. The bars in Figure 23 indicate how many participants there were in the data with the difference in median ratings specified on the x-axis.

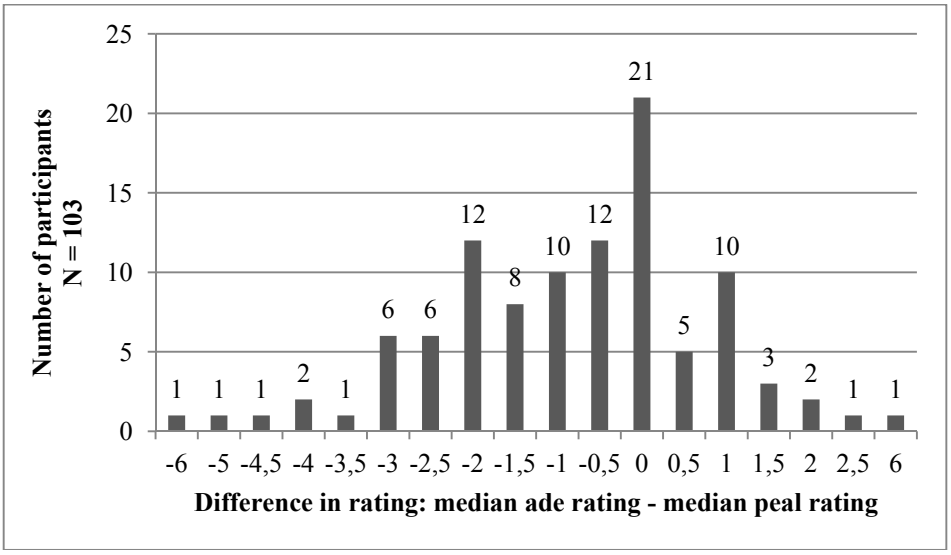


Figure 23. Individual difference in median ratings for adessive and *peal*-sentences in sentence rating task

From Figure 23 it can be seen that, in general, there is normal distribution – i.e. there are many participants for whom there was no difference. Averaged over all conditions, 21 participants out of 103 had the same median ratings for adessive and *peal*-sentences (the difference is 0); for 37 participants the difference was only -1, -0.5, 0.5 or 1. However, it can also be seen from Figure 23 that the left tail of the distribution is longer, i.e. on average the subjects prefer the adpositional construction – if we compare the number of participants for whom the difference was negative (i.e. for whom *peal*-sentences were more acceptable) with those for whom it was positive (i.e. those who rated adessive sentences higher), then the first group is considerably larger.

Another relevant issue is that of the average ratings for the specific four landmark-trajector pairs, i.e. items in the study. This is again an aspect that I will return to later when the mixed-effects model is implemented. For the time being suffice it to say here that all of the four landmark-trajector pairs (*table-vase, chair-bag, bed-blanket, couch-pillow*) received a median rating of 5.

5.3.6 Summary of the exploratory findings

Unlike with the results for the multivariate corpus study, we now have an overall difference in construction type. In the picture rating task, generalised over all conditions, the adessive and *peal*-sentences received a similarly high rating; in the sentence rating task, however, the *peal*-sentences receive, on average, a considerably higher rating than the adessive sentences. There seems to be an overall effect for all the three explanatory variables separately. In

addition to differences in the construction type, sentences with shorter Landmark phrases receive on average a higher rating than sentences with longer Landmark phrases. In addition, sentences with the word order sequence where the Landmark phrase comes first are preferred over sentences where the Landmark phrase follows the Trajector phrase.

As to the interactions between construction type and the two other variables, there seem to be none. If we take the interaction between construction type and length of the Landmark phrase, then for both constructions, shorter Landmark phrases are preferred over longer Landmark phrases; in both conditions, *peal*-sentences are rated higher than the adessive sentences. This result indicates that there seems to be no ground for prediction 1 that stated, based on the results of the multivariate corpus study, that adessive sentences with long Landmark phrases should receive higher ratings than the corresponding *peal*-sentences; however, the opposite seems to be true for the results of this rating task. Prediction 2 stated that adessive sentences with short Landmark phrases should receive lower ratings than the corresponding *peal*-sentences – this prediction does seem to hold.

The results for the combination of construction type and word order are similar to the interaction between construction type and length. Again, for both conditions – whether the Landmark comes first or second – *peal*-sentences are rated higher than the adessive sentences. In addition, for both constructions, sentences where the Landmark precedes the Trajector phrase are preferred over sentences where it follows the Trajector phrase. Prediction 3 which stated that adessive sentences with "lm1" should receive higher ratings than the corresponding *peal*-sentences does not seem to hold; prediction 4, on the other hand, does seem to hold – adessive sentences with "lm2" are rated lower than the corresponding *peal*-sentences. However, these results may simply be due to the fact that *peal*-sentences are, on average, rated higher than adessive sentences and the two additional factors – length of the Landmark phrase and word order – do not play a significant role.

The preliminary exploratory results described above are summarised in Table 84 where the values of the two construction types are cross-tabulated with the values of the other explanatory variables. Overall, there seems to be little evidence that either of the two variables in combination with type of construction play a role. Although the multivariate corpus study did produce a significant effect for length, it does not seem to be confirmed by this rating task. Naturally, we should not conclude that there is no effect, simply that the present experimental design was not "good" enough to find an effect if the effect is there. In order to check the validity of these exploratory findings, the next section presents a more sophisticated statistical analysis of the experimental data.

Table 84. Preliminary exploratory results of sentence rating task

	adessive	**peal 'on'**
Length of the Landmark phrase	short	short
Word order	lm1	lm1

5.3.7 Repeated measures ANOVA

In order to check if any of the differences highlighted above might not have arisen by chance and to test the specific predictions posited, this section presents a more sophisticated statistical analysis of the sentence rating task. Unlike the previous section on descriptive results, this section does credit to the true multivariate nature of the data set. As we saw above with the picture rating task, a number of different multivariate techniques can be used. In order to appreciate the consequences of using any of the different available techniques, the following section provides the analysis of the experimental data using all of these techniques. Although the inclusion of random effects in analysing the data is crucial, the results of using techniques that do not permit the inclusion of random effects are still included in order to demonstrate whether and what kind of differences there are in the results arising from this kind of omission. As we proceed from one technique to another, I will gradually build up a comparative table that allows us to compare the results arrived at by using the different techniques. Again, the purpose is to demonstrate that the choice of the multivariate technique for data analysis has its consequences when drawing conclusions about the results.

The repeated measures analysis of variance was carried out using IBM SPSS Statistics 19 software. A preliminary analysis of the participants' ratings was conducted using a construction (2) x version (2) repeated measures ANOVA. The analysis revealed a significant main effect of construction, $F(1, 101) = 33.65$, $p < 0.01$, partial $\eta^2 = 0.25$; there was no significant main effect of version [$F(1, 101) = 0.02$, $p = 0.89$, partial $\eta^2 = 0.00$] nor was the construction x version interaction significant [$F(1, 101) = 2.10$, $p = 0.41$, partial $\eta^2 = 0.01$] . For this reason, the results for the two versions were collapsed in all further analyses. Recall also that the experiment included 8 control sentences – 4 control sentences with a long Landmark phrase and 4 control sentences with a long Trajector phrase. Inclusion of these control sentences was necessary to exclude the influence of lexical effects, i.e. provided that the 8 control sentences all receive the same ratings and there are no significant differences between them, we can conclude that if the adessive sentences and *peal* sentences containing the long Landmark phrases differ, they differ in respect to length and not the specific lexical items included in them. The median ratings for all 8 control sentences was 7, the means were all close to 6 and the modes were 7. A one-way analysis of variance verified that the main effect for control sentences was not significant – $F(1, 816) = 1.18$, $p = 0.31$, partial $\eta^2 = 0.01$. However, if we

apply a repeated measures ANOVA, i.e. take the subjects as error terms, the results show that there is a borderline significant effect for control sentences – $F(1, 102) = 2.06$, $p = 0.05$, partial $\eta^2 = 0.02$; but as the value for partial η^2 indicates, the effect is very small. This result demonstrates the necessity of including subjects as error terms in the subsequent analysis, as some participants have a relatively low median and mean rating of the control sentences, which mirrors their overall tendency to assign low ratings.

Following the procedure described in Johnson (2008: 121–134) for a repeated measures analysis of variance, an ANOVA table is constructed with four experimental variables – construction, length of the Landmark phrase, word order, and subject (Table 85). In the first column we have the explanatory variables and their interactions for which we want to know whether they have an effect on the response variable of ratings. The second column lists the degrees of freedom (Df), the third one the sums of squares (Sum Sq), the fourth one gives the mean squares (Mean Sq). The last three columns in Table 85 are of main interest – from there we get the F-values and *p*-values, which we use to decide whether a factor or an interaction has a significant effect on the response variable or not. The last column gives the measure for the effect size – partial eta-squared values.

Table 85. Results for construction (2) x length (2) x word order (2) repeated measures ANOVA (sentence rating task)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	η_p^2
construction (cx)	1	105.08	105.077	33.758	<0.001	0.23
cx:subject	102	317.49	3.113			
length	1	200.78	200.783	138.23	<0.001	0.58
length:subject	102	148.16	1.453			
word order (wo)	1	9.028	9.0280	10.0015	<0.001	0.09
wo:subject	102	91.949	0.9015			
cx:length	1	5.820	5.8199	8.4127	0.005	0.08
cx:length:subject	102	70.563	0.6918			
cx:wo	1	0.073	0.07289	0.1218	0.7278	0.001
cx:wo:subject	102	61.029	0.59832			
length:wo	1	69.177	69.177	62.277	<0.001	0.47
length:wo:subject	102	113.300	1.111			
cx:length:wo	1	0.182	0.18211	0.4465	0.5055	0.004
cx:length:wo: subject	102	41.607	0.40791			

From Table 85 it can be seen that there are three significant main effects: moderate significant effect for construction, $F(1, 102) = 33.76$, $p < 0.001$, partial $\eta^2 = 0.23$; large effect for length of the Landmark phrase, $F(1, 102) = 138.23$, $p < 0.001$, partial $\eta^2 = 0.58$; and a very small effect for word order, $F(1, 102) = 10.02$, $p < 0.001$, partial $\eta^2 = 0.09$. Out of the three two-way interactions

two were significant: a very small effect for construction x length, $F(1, 102) = 8.41$, $p = 0.005$, partial $\eta^2 = 0.08$ and a large effect for length x word order, $F(1,102) = 62.28$, $p < 0.001$, partial $\eta^2 = 0.47$. The two-way construction x word order interaction and the three-way construction x length x word order interaction were not significant. Compared to the descriptive results above, one result is especially noteworthy – the interaction between construction type and length has become significant, although with a very small effect size. Table 81 and Figure 19 above, however, show that there is an additive effect instead. This anomaly in the results may be produced due to the fact that we are treating the response variable as a measurement on interval scale and that we have now factored in subjects as error terms in the analysis.

We can inspect the findings further by looking at the possible pairwise comparisons, because at the moment we only know that there is a significant effect, but we have no idea in which direction it goes. In order to do that, we can compare the mean ratings that were provided in the descriptive tables in the previous section. Looking at the effect of construction, we already saw in the previous section that there was indeed a significant difference between the average ratings – the adessive sentences (mean = 4.34, SD = 2.00) were rated considerably lower than the *peal*-sentences (mean = 5.05, SD = 1.85); this was confirmed with a paired *t*-test and we also have confirmation for this result from repeated measures ANOVA. We also mentioned in the previous section that the sentences with short Landmark phrases received higher ratings than sentences with long Landmark phrases: the mean acceptability rating for the former was 5.19 (SD = 1.82) and for the latter 4.20 (SD = 1.96). Again, the results reported in Table 85 confirm that this difference is significant. The repeated measures ANOVA also confirms that the difference between the mean ratings for the two word order sequences differ significantly – the average mean rating was higher for the "lm1" sequence (mean = 4.80, SD = 2.00) than for landmark second, "lm2", sequence (mean = 4.59, SD = 1.91).

Furthermore, we can also inspect the significant two-way interactions further. According to Johnson (2008: 128), the usual strategy in psycholinguistics is to perform another repeated measures ANOVA on a subset of the data. Let us first look at length x construction. The results of a repeated measures ANOVA on the subset of data including only data points for long Landmark phrases reveals a significant effect of construction ($F(1, 102) = 40.09$, $p < 0.001$, partial $\eta^2 = 0.28$). This result confirms that the difference in mean ratings we saw in Table 81 above is significant – *peal*-sentences received higher ratings (mean = 4.64, SD = 1.91) than adessive sentences (mean = 3.76, SD = 1.92). Another repeated measures ANOVA was performed on the subset of short Landmark phrases. The effect of the construction is again significant ($F(1, 102) = 17.02$, $p < 0.001$, partial $\eta^2 = 0.14$) – *peal*-sentences were rated higher than the adessive sentences with short Landmark phrases as well. The respective means were 5.46 (SD = 1.67) and 4.91 (SD = 1.91). Based on the partial eta-squared value we can conclude that the difference is more prominent with long Landmark phrases. However, we should keep in mind that the overall

effect size for the interaction between construction and length of Landmark phrase was very small (partial $\eta^2 = 0.08$) compared to that of length alone (partial $\eta^2 = 0.58$) and the length x word order interaction (partial $\eta^2 = 0.47$). We will turn to this interaction now.

After running two separate analyses on the subsets of long [($F(1, 102) = 17.34$, $p < 0.001$, partial $\eta^2 = 0.28$] and short Landmark phrases [$F(1, 102) = 53.48$, $p < 0.001$, partial $\eta^2 = 0.14$], it becomes evident that there is a significant difference between the two word order sequences; the effect is small for short Landmarks and medium for long Landmarks. While the mean rating for sentences where the Landmark comes second do not differ considerably for short (4.38, SD = 1.93) and long Landmark phrases (4.79, SD = 1.86), there is a considerable difference in the mean ratings for sentences where the Landmark comes first and the Landmark phrase is short (5.58, SD = 1.69) and long (4.01, SD = 1.98) – the former are rated much higher than the latter. A possible explanation for these results takes into account the general information flow – since we are usually interested in the location of the Trajector, we tend to place the Landmark phrase second, irrespectively whether it is long or short. When the Landmark is short, however, it can come at the beginning of the sentences, since the longer Trajector phrase comes at the end (cf. the principle of end-weight: Hawkins 1994, 2004; Wasow 1997, 2002).

Table 86 summarises the results for a three-way repeated measures ANOVA.

Table 86. Comparison of the multivariate techniques used for the analysis of sentence rating task (rep. measures ANOVA)

Stat. method	rep. measures ANOVA	
	sign.	effect
construction (cx)	sign.	0.23
length of Lm phrase (length)	sign.	0.58
word order (wo)	sign.	0.09
cx*length	sign.	0.08
cx*wo	not sign.	0.001
length*wo	sign.	0.47
cx*length*wo	not sign.	0.004
random-effect	yes	
response variable	interval	

As was seen with the exploratory results, there is a significant main effect for the variables construction type and length of the Landmark phrase – *peal*-sentences are rated higher than adessive sentences and sentences with short Landmark phrases higher than long Landmark phrases. Although there is a

significant effect for word order (Landmarks are preferred to come first), the effect itself is very small. There is no effect for the interaction between construction type and word order, and although the interaction between construction type and length is significant, the effect size is very small. There is, however, a very large significant effect for the interaction between length and word order. The three-way interaction between all of the explanatory variables is not significant. All in all, the results seem to point towards the conclusion, which was already voiced above – the present experimental design gives little confidence for claiming that the combination of the explanatory variables (most importantly construction type and length, and construction type and word order) plays a role. Instead, there is a very strong main effect for length of the Landmark phrase and its interaction with word order, construction type does not enter the picture, save for the very general observation the *peal*-sentences are rated higher than the adessive sentences. I will continue the analysis with linear mixed-effects regression, which still treats the response variable as interval, but allows an extra random effect to be taken into account besides subjects – the specific sentences used in the rating task, i.e. the item-based random effects.

5.3.8 Linear mixed-effects regression

The research question we are asking with regression modelling is whether a rating of a sentence in the experiment can be predicted from the locative construction, length of Landmark, and the specific word order sequence used in this sentence. See the discussion in the previous section on some of the arguments why mixed-effects models may be considered more useful than, for example, analysis of variance. The analysis itself is carried out in R following the procedure described in Baayen (2008: 242–259). In this statistical procedure the ratings on the scale from 1 to 7 are modelled as depending on the predictor variables of construction type, length of the Landmark phrase and word order. In the formula, we also have random-effects terms for Subjects and Items (in this case, the individual sentences used in the questionnaire). Including Subjects and Items as random effects in the model allows it to take into account that some subjects provide higher ratings and some lower ratings. Similarly, some sentences (or items, as they are referred to here), may be rated lower and we thus may want to adjust the intercept for the individual sentences by means of a random-effects term as well (Baayen 2008: 245). In R, one can inspect the actual adjustments made for specific subjects and specific items with the `ranef()` function (Baayen 2008:246).

First, a full model with all three explanatory variables and all interactions (including the three-way interaction) was built. Since, however, the three-way interaction between the three explanatory variables and the interaction between type of Landmark and word order turned out to be insignificant, a new model was built excluding these interactions. A new model was built with all the three explanatory variables and the interactions between construction and type of

Landmark, and construction and word order. Since the results show that the three-way interaction and the two-way interactions between construction type and word order, and construction type and length are not significant, they are taken out of the model. The object for the final linear mixed-effects model is presented in Figure 24 with construction, length of Landmark, and word order as fixed-effects and Subjects and Items as random-effects. In the summary of the model object in Figure 24, after a list of summary statistics that describe the quality of the fit of the model to the data, there is a table with the random effects in the model, followed by a table with fixed effects; the summary concludes with a table of correlations of the fixed effects. The part of the summary dealing with fixed effects gives a table of coefficients of the fixed effects – in this case the coefficient for the intercept and for the slope of construction (CX), length of Landmark (LM), word order (WO) and their interactions. The table also lists the associated standard errors, *t*-values, and the *p*-values on the *t*-statistic.

```
Linear mixed model fit by REML
Formula: RATING ~ CX + LENGTH + WO + LENGTH * WO + (1 |
SUBJECT) + (1 |      SENTENCE)
   Data: morfall
   AIC   BIC logLik deviance REMLdev
 12977 13026  -6481    12951   12961
Random effects:
 Groups   Name        Variance Std.Dev.
 SUBJECT  (Intercept) 0.52085  0.7217
 SENTENCE (Intercept) 0.09891  0.3145
 Residual             2.76557  1.6630
Number of obs: 3296, groups: SUBJECT, 103; SENTENCE, 32

Fixed effects:
                   Estimate Std. Error t value   Pr(>|t|)
(Intercept)          3.6575     0.1572  23.261   0.000
CXpeal               0.7142     0.1254   5.695   0.000
LENGTHshort          1.5667     0.1773   8.834   0.000
WOlm2                0.3701     0.1773   2.087   0.037
LENGTHshort:WOlm2   -1.1590     0.2508  -4.621   0.000
```

Figure 24. Output for the linear mixed-effects model (sentence rating task)

Before moving on to interpreting the results, the residuals were checked for potential problems with the model specification. It seems that the model is not coping very well with very high ratings. This means that the model does not fit very well and we have to be careful in interpreting the results. As noted above with the linear mixed-effects regression for picture rating task, the summary tables for mixed-effects models obtained with `lmer()` do not list the pro-

portion of variance (R^2) accounted for, a statistic that is commonly used to assess the goodness-of-fit of the model. This is because there are a number of different sources of variance that are modelled jointly; in addition to the variance explained by fixed effects, we have the variance explained by the by-subject and by-item random effects (Baayen 2008: 258).

Bearing in mind that the model we see in Figure 24 may not fit the data well, we can take a look at the results. As noted above, in the model with interactions, everything is recalibrated and the main effects by themselves are no longer very informative. However, Figure 24 does confirm what we saw above with the exploratory findings and with the analysis of variance – *peal*-sentences received significantly higher ratings than adessive sentences. The way we can see this from Figure 24 is that the estimate for the predictor construction (CX) with the level *peal* (CXpeal) given in the second column in the summary output for fixed effects is positive (0.7141), indicating scores are predicted to be higher; this result is significant (t-value = 5.695 and $p < 0.001$). Similarly, in accordance with the exploratory results and the results obtained with ANOVA, sentences with short Landmark phrases were rated higher than sentences with long Landmark phrases – this is again demonstrated in Figure 24, where we have a positive estimate (1.5667) for the short Landmark phrases (LENGTHshort). The t-value (8.83) and p-value (< 0.001) confirm that this result is statistically significant. The fact that this predictor is actually the strongest, as seen from above in the analyses of variance where it had the biggest effect size, is also seen in the mixed-effect model technique – it has the largest value for its estimate.

According to the model presented in Figure 24, word order is also a significant predictor (t-value = 2.09, $p = 0.04$). The mixed-effect modelling technique also reaches the same conclusion as the ANOVAs above concerning the interaction between length of the Landmark phrase and word order. The coefficient for short Landmark phrases (LENGTHshort) with the landmark second word order sequence (WOLm2) has a negative coefficient (LENGTHshort:WOLM2 = -1.1590) indicating that this combination received significantly (t-value = -4.62, $p < 0.001$) lower ratings.

The summary of the results of linear mixed-effects regression are given in Table 87, where the results are compared with a repeated measures ANOVA.

Table 87. Comparison of the multivariate techniques used for the analysis of sentence rating task (rep. measures ANOVA vs. linear mixed-effects reg.)

Stat. method	rep. measures ANOVA		linear mixed-effects regression
	sign.	effect	
construction (cx)	sign.	0.23	sign.
length of Lm phrase (length)	sign.	0.58	sign.
word order (wo)	sign.	0.09	sign.
cx*length	sign.	0.08	not sign.
cx*wo	not sign.	0.001	not sign.
length*wo	sign.	0.47	sign.
cx*length*wo	not sign.	0.004	not sign.
random-effect	yes		yes
response variable	interval		interval

The results show that both techniques indicate that there is a significant main effect of all three explanatory variables. Both techniques also converge on the result that the interaction between length of the Landmark phrase and word order is significant, but that the interaction between construction type and word order and the three-way interaction between all three variables are not significant. There is diverging evidence as pertains to the interaction between construction type and length. A three-way repeated measures ANOVA suggests that there is an interaction, although the effect is very small, yet linear mixed-effects regression demonstrates that this interaction is not significant. This may, again, point towards the general conclusion that ANOVA is more prone to picking up differences when in fact they are not present. In addition, the linear mixed-effects regression model accounted for an extra source of error – item-based random effects. At the same time, as was pointed out before, the linear mixed-effects regression model has some problems with the model fit. Hopefully, the use of ordinal logistic regression, both simple and mixed-effects, will shed some more light on the results.

5.3.9 Ordinal logistic regression

For the purposes of comparison, I will first present the data analysis where the issue of random-effects for subjects is ignored and then use a new method that allows the inclusion of random-effects also in ordinal logistic regression modelling[72]. The first follows the procedure described in Norušis (2011: 69–89) and the analysis itself will be carried out in SPSS; the second follows the procedure described in Divjak (in progress) and the analysis is carried out in R. With ordinal logistic regression, or the proportional odds model as it is also

[72] Currently, only a single random term is allowed in the model.

known, the research question is similar as with linear regression and can be stated as whether a rating given to a sentence can be predicted from the combination of factors: locative construction, length of the Landmark phrase and word order.

The first step in model building is fitting the full model with rating as the response variable and construction type, length of the Landmark phrase, word order and their interactions as factors or predictor variables. However, since the three-way interaction between the explanatory variables, as well as the two-way interaction between construction type and word order, are not significant, they are left out of the model. Before interpreting the results, the assumption of the parallel lines or the parallel odds assumption, and other model criticism needs to be checked. The model assumption is that the regression coefficients are the same for all five categories. Since the observed significance level in Table 88 is large, there is not sufficient evidence to reject the parallelism hypothesis, i.e. the model does not violate the assumption.

Table 88. Test of parallel lines for ordinal logistic regression model (sentence rating task)

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	293.638			
General	279.015	14.624	25	.950

In assessing the model fit for ordinal logistic regression models in SPSS, various statistics are available (see section 3.4.1 above). One way of assessing whether the model fits well, is by looking at the difference between observed cell counts and expected cell counts. Table 2 in Appendix 3 presents the observed and expected (given in brackets) cell counts for all the experimental conditions by the 7 rating categories. There does not seem to be much difference between the observed and expected counts. The Pearson and Deviance goodness-of-fit statistics given in Table 89 confirm this – the observed and expected cell counts are similar and we can conclude that the model fits well.

Table 89. Goodness-of-fit statistics for ordinal logistic regression model (sentence rating task)

	Chi-Square	df	Sig.
Pearson	33.407	37	.638
Deviance	33.751	37	.622

Continuing with the model criticism, one should also pay attention to the strength of association. According to Norušis (2011: 81) there are several R^2-like statistics that can be used to measure the strength of the association

between the response variable and the predictor variables. But Norušis (2011: 81) also points out that they "are not as useful as the R^2 statistic in regression, since their interpretation is not straightforward". Baayen (2008: 204) reaches a similar conclusion. For the experimental data at hand, the values of the pseudo *R*-square statistics are small as can be seen from Table 90. However, they are higher than the corresponding values for ordinal logistic regression model for picture rating task. Still, the model may not be very strong as the association is not that strong.

Table 90. Pseudo R-square statistics for the ordinal logistic regression model (sentence rating task)

Cox and Snell	.126
Nagelkerke	.129
McFadden	.036

Considering that the model diagnostic figures are within the acceptable range, it can be assumed that the predictions made by the model can be trusted and we can proceed with looking at the actual results. The output of the analysis done in SPSS is presented in Table 91. The output gives the estimated coefficients for the model. As is the case with categorical predictors in models with intercepts, the number of coefficients displayed is one less than the number of categories of the variable and in SPSS the reference category is the last category (cf. R where the reference category is the first category). Larger coefficients indicate an association with larger scores; alternatively, negative coefficients indicate that lower scores are more likely. The Wald statistic is the square of the ratio of the coefficient to its standard error. Small significance level indicates that the null hypothesis that the ratio is zero can be rejected.

Table 91. Output for the ordinal logistic regression model (sentence rating task)

	Estimate	Std. Error	Wald	Df	Sig.
[RATING=1]	-3.036	.099	931.48	1	.000
[RATING=2]	-1.935	.085	515.29	1	.000
[RATING=3]	-1.265	.081	244.56	1	.000
[RATING=4]	-.697	.079	78.42	1	.000
[RATING=5]	-.034	.078	.19	1	.664
[RATING=6]	.945	.080	139.67	1	.000
[CX=ade]	-.525	.088	35.25	1	.000
[LENGTH=long]	-.228	.107	4.58	1	.032
[WO=lm1]	.836	.089	88.28	1	.000
[CX=ade]*[LENGTH=long]	-.288	.124	5.44	1	.020
[LENGTH=long]*[WO=lm1]	-1.185	.125	90.53	1	.000

There appears to be a relationship between the explanatory variables construction type, length of Landmark, word order and the response variable rating. As was noted above, since there are significant interactions in which the main effects are involved, the interpretation of the main effects is not straightforward. The results show that there is a relationship between the interaction of construction type with length of the Landmark and rating. For any rating level, adessive sentences with long Landmark phrases ([CX=ade]*[LENGTH=long]) receive a lower rating (estimate is negative: -.288); this result is significant (p = .020). There is also relationship between the interaction of length of the Landmark phrase and word order. For any rating level, sentences with long Landmark phrases with the word order "lm1" ([LENGTH=long]*[WO=lm1]) receive a lower rating (the estimate is negative: -1.185); this result is again significant ($p < .001$).

The results of ordinal logistic regression are summarised in Table 92, where the results are compared with a repeated measures ANOVA and linear mixed-effect regression.

Table 92. Comparison of the multivariate techniques used for the analysis of sentence rating task (rep. measures ANOVA vs. linear mixed-effects reg. vs. ordinal log. reg.)

Stat. method	rep. measures ANOVA		linear mixed-effects regression	ordinal logistic regression
	sign.	effect		
construction (cx)	sign.	0.23	sign.	sign.
length of Lm phrase (length)	sign.	0.58	sign.	sign.
word order (wo)	sign.	0.09	sign.	sign.
cx*length	sign.	0.08	not sign.	sign.
cx*wo	not sign.	0.001	not sign.	not sign.
length*wo	sign.	0.47	sign.	sign.
cx*length*wo	not sign.	0.004	not sign.	not sign.
random-effect	yes		yes	no
response variable	interval		interval	ordinal

All three techniques converge on the following results: there is a significant main effect of all three explanatory variables, the two-way interaction between length of the Landmark phrase and word order is significant, but the two-way interaction between construction type and word order and the three-way interaction are not significant. There is diverging evidence on the two-way interactions of construction type x length of the Landmark phrase. The ordinal logistic regression model confirms the results of the repeated measures ANOVA that indicated the interaction between construction type and length is significant, although linear mixed-effects model did not replicate this result.

5.3.10 Ordinal mixed-effects logistic regression

As was mentioned above for the picture rating task, this specific method is highly desirable for the kinds of experimental data discussed here, but since it is fairly new and the R package `ordinal` for conducting this analysis came out only in 2010, there are to date very few practical introductions to using this method. The only available document is the R manual (available at: http://cran.r-project.org/web/packages/ordinal/ordinal.pdf) and the only linguistic study that uses this method is to my knowledge Divjak (in progress). As the technique is still new, there are as yet not many model criticism measures available. The very important advantages of this technique are that it, first and foremost, treats the ratings as ordinal and secondly, it allows the inclusion of random-effects.

The model was built stepwise, starting with the most complex model that included all the predictors (construction type, length of the Landmark phrase, and word order) and their interactions. Since the two-way interaction between construction type and word order, and the three-way interaction is not significant, these were taken out of the model. The best model (the output of which is given in Figure 25) contains the random effect for Subject and as fixed effects the three predictors and the interaction between construction type and length of the Landmark phrase and the interaction between length of the Landmark phrase and word order. One diagnostics[73] that the `ordinal` package in R currently does provide is the condition number of the Hessian measuring the empirical indentifiability of the model (Divjak in progress). The condition number of Hessian for this model equals 252.3197, signalling that the model predicts the ratings well using the combination of explanatory variables included in the model.

The three main effects in the model (Figure 25) are significant – construction type (CX), length of the Landmark phrase (LENGTH) and word order (WO). Importantly, the interaction between construction and length of the Landmark is also significant, as is the interaction between length of the Landmark phrase and word order. For any rating level, *peal*-sentences with short Landmark phrases (CXpeal:LENGTHshort) receive a lower rating (estimate is negative: -.2943); this result is significant (p = .018). For any rating level, sentences with short Landmark phrases with the word order "lm2" (LENGTHshort:WOlm2) receive lower ratings (the estimate is negative: -1.2811); this result is again significant (p < .001).

[73] There are other diagnostics available, but since they are not implemented in the package itself, they need to be calculated manually.

```
clmm(location = as.factor(RATING) ~ CX + LENGTH + WO + CX *
LENGTH + LENGTH * WO, random = as.factor(SUBJECT), data =
morfall, Hess = TRUE, link = "logistic")

Random effects:
                          Var   Std.Dev
as.factor(SUBJECT) 0.5770887 0.7596636

Location coefficients:
                   Estimate Std. Error z value  Pr(>|z|)
CXpeal               0.9001   0.0880    10.2317 < 2.22e-16
LENGTHshort          1.8439   0.1128    16.3452 < 2.22e-16
WOlm2                0.3575   0.0873     4.0963 4.1972e-05
CXpeal:LENGTHshort  -0.2943   0.1248    -2.3580 0.018376
LENGTHshort:WOlm2   -1.2811   0.1265   -10.1262 < 2.22e-16

No scale coefficients

Threshold coefficients:
    Estimate Std. Error z value
1|2  -1.8166   0.1191   -15.2464
2|3  -0.6308   0.1098    -5.7456
3|4   0.1133   0.1088     1.0419
4|5   0.7481   0.1097     6.8223
5|6   1.4888   0.1119    13.3086
6|7   2.5735   0.1170    22.0017

log-likelihood: -5806.27
AIC: 11636.54
Condition number of Hessian: 252.3197
```

Figure 25. Output for the Cumulative Link Mixed Model (sentence rating task)

5.4 Discussion of the results: sentence rating task

The overview of the results of all four multivariate techniques used to analyse the results of the sentence rating task are given in Table 93. The results that are the same across all techniques are in bold.

Table 93. Comparison of the multivariate techniques used for the analysis of sentence rating task

Stat. method	rep. measures ANOVA		linear mixed-effects regression	ordinal logistic regression	ordinal mixed-effects logistic regression
	sign.	effect			
construction (cx)	**sign.**	**0.23**	**sign.**	**sign.**	**sign.**
length of Lm (length)	**sign.**	**0.58**	**sign.**	**sign.**	**sign.**
word order (wo)	**sign.**	**0.09**	**sign.**	**sign.**	**sign.**
cx*length	sign.	0.08	not sign.	sign.	sign.
cx*wo	not sign.	0.001	not sign.	not sign.	not sign.
length*wo	**sign.**	**0.47**	**sign.**	**sign.**	**sign.**
cx*length*wo	not sign.	0.004	not sign.	not sign.	not sign.
random-effect	yes		yes	no	yes
response variable	interval		interval	ordinal	ordinal

All four techniques indicate that there is a main effect of all three explanatory variables – construction type, length of the Landmark phrase, and word order. In addition, all models also indicate that the two-way interaction between length of the Landmark phrase and word order is significant and that the two-way interaction between construction type and word order, and the three-way interaction do not have a significant effect. There is, however, diverging evidence for the two-way interactions between construction type and length. The interaction is significant in the repeated measures ANOVA, ordinal logistic regression model, and in the ordinal mixed-effects logistic regression model; it is not significant, in the linear mixed-effects regression model. The difference here may be due to the fact that the latter model is the only model that includes two error terms – it includes both subjects and items as random effects. The repeated measures ANOVA confirms that even if there is an effect, the size is very small (partial η^2 is only 0.08).

In general, the unfortunate conclusion to be drawn is that the four predictions at the beginning of this chapter found no verification. It was predicted on the basis of the multivariate corpus study that the combinations between construction type and length of the Landmark on the hand, and construction type and word order on the other hand, have a significant effect. As the

exploratory findings indicated, there seems to be no interaction and the results are additive instead. That means that the *peal*-sentences are rated significantly higher than the adessive sentences irrespective of the experimental condition. The multivariate techniques confirm this. None of the models produced a significant effect for the interaction between construction type and word order. This, in itself is a nice result in respect that it mirrors the results of the multivariate corpus study where the relative position of the Landmark phrase was not found to be significant either. (But cf. the results of the picture rating task where there is a significant effect for this interaction; I will take up this point in the next chapter where I provide a more detailed discussion of the results.) Although some of the models did produce a significant effect for the interaction between construction type and length of the Landmark phrase, there is diverging evidence, and if the effect is there, it is very small. Furthermore, if we look at the results, it becomes clear that the prediction that the adessive is preferred with longer Landmark phrases, something that the results of the multivariate corpus study do show, is not confirmed. If anything, the situation is the opposite – *peal*-sentences with longer Landmark phrases received a higher rating.

Still, we should be careful in concluding that there is no effect for the combination of construction type and length, and construction type and word order – the fact that I did not find a significant effect may have simply arisen from the limitations of the experimental design. One of the reasons why *peal*-sentences received a significantly higher rating in all experimental conditions may have to do with the specific set of stimuli. The stimuli included lexemes that denoted pieces of furniture. As explained above in section 3.2.2, these lexemes were selected to keep the semantic factor of type of Landmark constant; another criterion for selecting these Landmarks was that they are sufficiently frequently used with both constructions. As the picture rating task shows, pieces of furniture tend to favour the adpositional construction over the case construction; hence, semantics may still have played a role in the results of the sentence rating task. A further experiment is desirable with a different set of stimuli and possible a different experimental design to investigate the issue of length further.

6. DISCUSSION AND CONCLUDING REMARKS

The last two chapters presented the results of the four studies included in the dissertation. Chapter 4 reported the results of the two corpus studies and Chapter 5 the results of the two rating tasks. This chapter discusses these results from the perspective of converging and diverging evidence and provides some concluding remarks. First, it looks at the results within the dissertation, i.e. whether and how the experimental results mirror those of the two corpus studies (section 6.1). It also makes reference to Chapter 1, where the importance of methodological pluralism was discussed and stressed. The second part of the chapter, section 6.2, takes a wider look at the results and compares the corpus and experimental studies reported here with other studies on grammatical synonymy. Section 6.3 assesses the usefulness of the specific methodologies used in the dissertation and discusses whether the same methodological approach is practical in future research on grammatical synonymy. The chapter ends with a list of possible suggestions for further corpus and experimental studies to be undertaken within this research domain and some general conclusions.

6.1 Evidence in Linguistics: Comparison of corpus and experimental results

The aim of the dissertation is to provide a multivariate account on the alternation between the adessive case and the adposition *peal* 'on' in Estonian – an instance of grammatical synonymy which has not been previously studied from a detailed theoretical and empirical perspective. The dissertation looks at eleven semantic (type of relation between Trajector and Landmark, type, animacy, number, and mobility of Landmark and Trajector, relative size of the Trajector in relation to the Landmark, and verb group) and nine morphosyntactic variables (length, complexity, and syntactic function of the Landmark phrase, word class of Trajector and Landmark, case form of Trajector, clause type, position of the Landmark phrase, the relative position of Trajector and Landmark). These variables were selected on the basis of the results of the few previous studies on the alternation between cases and adpositions in other Finno-Ugric languages (Bartens 1978, Ojutkangas 2008), other alternation phenomena discussed in the literature (e.g. the numerous studies on the English particle placement, genitive and dative alternation), and on the basis of my own native speaker intuition.

The first two studies are corpus studies, while the latter two are experimental studies that test some of the hypotheses posited on the basis of corpus studies and my own intuition as a native speaker of Estonian. The first corpus study is multivariate in nature and the aim was to look at all the 20 variables simultaneously, providing both univariate (chi-squared tests, *t*-tests) and multivariate results (binary logistic regression). By employing regression analysis, it was

possible to determine the contribution of the different semantic and morphosyntactic variables to the alternation and calculate the relative strength of each individual variable. The second corpus study focuses purely on the semantic variable type of Landmark and employs the methodology of distinctive collexeme analysis. This method was used in order to see if there are specific lexemes that distinguish between the two constructions. The two rating tasks (the picture rating task and the sentence rating task) test the validity of three hypotheses that are proposed on the basis of the corpus data, the studies on other alternation phenomena, and my own intuition as a native speaker of Estonian.

The picture rating task looks at the semantic variable type of Landmark and the morphosyntactic variable word order. These variables were selected to validate the claim that, first of all, the case construction is used with larger, static places (e.g. *turg* 'market') and the adpositional construction with smaller, mobile things (e.g. *kapp* 'wardrobe'), and secondly that the adessive construction is preferred when the Landmark precedes the Trajector (e.g. *laual on vaas* 'on the table is a vase') and the adpositional construction when it follows the Trajector (e.g. *vaas on laual* 'the vase is on the table'). Since the two constructions express spatial relations, it was felt necessary to provide visual stimuli. The sentence rating task looks at two morphosyntactic variables – length of the Landmark phrase and the relative position of Trajector and Landmark. The reason why specifically these variables were selected has to do with converging evidence. The multivariate corpus study shows that length of the Landmark phrase is one of the strongest predictors – the longer the Landmark phrase, the more probable it is that the adessive construction is used. It was decided to validate this result in a sentence rating task. The picture rating task provided conflicting results as pertains to the variable relative position of Trajector and Landmark. The multivariate corpus results indicate that there is a slight preference for the adpositional construction to be used when the Landmark follows the Trajector – this is also my intuition as a native speaker of Estonian. However, the picture rating task leads to a different conclusion – the adessive received higher ratings when the Landmark followed the Trajector than the corresponding sentences with the adpositional construction. It was therefore decided to include this variable in the sentence rating task as well.

The four studies reported in the dissertation show that the semantic variables type and mobility of the Landmark, the relative size of the Trajector in relation to the Landmark, and verb group play a role. From among the nine morphosyntactic variables examined, the following five are significant: length and morphological complexity of the Landmark phrase, word class of Trajector and Landmark, and, to some extent, word order. The variables that did not play a significant role include the following: the various semantic properties of the Trajector, case form of Trajector, animacy and number of Landmark, clause type, type of relation between Trajector and Landmark, and the position of the Landmark phrase. It is important to bear in mind that I am not claiming that these variables do not play a role at all. It may be that these variables become

significant once other variables are included and another logistic regression model is built. It is always important to look at how the different variables behave when taken together. For example, the univariate corpus results show that clause type, syntactic function of the Landmark phrase, number of Landmark and Trajector, and type of relation between Trajector and Landmark are significant (although the effect size is very small), but the multivariate logistic regression does not confirm this. On the other hand, the univariate results suggest that there is no significant association between type of construction and relative position of Trajector and Landmark, but once we take all the variables into account simultaneously, the word order variable becomes significant. These results demonstrate the limitations of employing only univariate statistical techniques and illustrate the usefulness of such multivariate techniques as logistic regression.

The following general guidelines can be given as to when one construction is preferred over the other construction by native speakers of Estonian in present-day written Estonian:

i) The **adessive** construction tends to be used with:
   - morphologically complex and long noun phrases as Landmarks (e.g. *kirjutuslaud* 'writing-desk');
   - static places as Landmarks (e.g. *turg* 'market');
   - Landmarks that are bigger than Trajectors;

ii) The ***peal***-construction tends to be used with:
   - short and simple noun phrases, especially with pronouns (e.g. *see* 'this');
   - small, mobile things as Landmarks (e.g. *kapp* 'wardrobe');
   - Landmarks that are of the same size as Trajectors;
   - verbs of existence (e.g. *vaas on laua peal* 'the vase is on the table').

In general, the two corpus studies and the picture rating task provide converging evidence and suggest that the semantic variables type and mobility of Landmark play a role in the alternation between the adessive and the adposition *peal* 'on'. The two locative constructions in Estonian exhibit a division of labour: the case construction is used with larger, static places, while the adpositional construction is preferred with smaller, mobile things. This confirms the results of Bartens (1978) and Ojutkangas (2008) who found a similar result for the alternation between interior locative cases and the corresponding adpositions in the Saami and Finnish languages respectively. The results reported in this dissertation also confirm the general prediction made in the literature that cases are more abstract and express more frequent spatial relations than adpositions (Comrie 1986, Luraghi 1991: 66–67, Hagège 2010: 37–38, and Lestrade 2010). Although the experimental results show a strong significance of the interaction between type of construction and type of Landmark, the corpus results show

that this result is by no means categorical. The semantic variables type and mobility of Landmark exhibit only a moderate effect size in the multivariate corpus analysis compared to such morphosyntactic variables as length or morphological complexity. Both the multivariate corpus study, as well as the distinctive collexeme analysis indicate that the adpositional construction can be (and is) used with larger, static places as well, while the adessive construction may very well be used with smaller, manipulable things. Nevertheless, the general tendency is there and as the experimental results of the picture rating task indicate, people are sensitive to this semantic division between the two constructions in the conceptual space.

The results of the multivariate corpus study demonstrate that the verbal predicate plays a role in the alternation between the two locative constructions in Estonian. Although the significance of this variable was not tested in an experimental task, the binary logistic regression predicts that verbs of existence (especially the copula *olema* 'be') prefer the adpositional construction. The reasons for this are manifold and future research clearly needs to employ an adequate experimental design to confirm which of the possible explanations hold. One possible explanation why the copula *olema* 'be' is used with *peal* 'on' and not the adessive is to avoid ambiguity. Adessive used together with the copula *olema* 'be' is frequently used in the Estonian possessive constructions and language users, in order to avoid the ambiguity between the locative and possessive functions of the adessive, prefer to use the adpositional construction with the verb *olema* 'be'. Another possible explanation relates to the preference of the *peal*-construction to occur in shorter and less complex environments. The present and past forms of the verb *olema* 'be' are relatively short in Estonian and hence extra explicitness is expressed with the adpositional construction.

As for the morphosyntactic variables that were examined in both corpus as well as experimental studies, the dissertation provides intriguing results. The multivariate corpus study indicates that the length of the Landmark phrase is one of the strongest predictors in the alternation between the adessive and the adposition *peal* 'on'. The adessive construction is preferred with longer and morphologically more complex Landmark phrases and the adpositional construction, on the other hand, in shorter and less complex environments. The motivation for this division of labour comes from the related principles of iconicity and economy (cf. Haiman 1983). On the one hand, language users do not wish to make an already long and complex phrase longer by adding an extra word to it and prefer the synthetic case inflection instead. On the other hand, they do wish to make a short and fairly opaque phrase more explicit by adding an extra word – *peal* 'on'. It is interesting that the sentence rating task does not confirm this prediction. The results demonstrate that for the specific set of stimuli, the adpositional construction is preferred with both short and long Landmark phrases. This divergence of evidence may point towards two possible lines of argumentation. One possible explanation is that the rating task only included one type of Landmarks – pieces of furniture. The aim was to control for the semantic variables and it was intentionally decided to include only

pieces of furniture, because the results of both corpus studies confirm that with pieces of furniture, the two constructions are used with more or less equal frequency. However, the sentence rating task may mirror the results of the picture rating task, where it was shown that participants rated *peal*-sentences with pieces of furniture slightly higher than the corresponding adessive sentences. Hence, it may be the case that semantics overrides morphosyntax and that the type of Landmark is a more significant variable than length when language users choose between the two constructions. It would be interesting to see how and if length plays a role when other types of Landmarks are included in the experimental design.

Another set of intriguing results pertains to the morphosyntactic variable relative position of Trajector and Landmark. The univariate results of the first corpus study demonstrate that there is no significant association between the relative position and the type of construction – both the adessive and the *peal*-constructions are frequently used in the middle or at the end of the clause and both prefer the Landmark to follow the Trajector (e.g. *vaas on laual/laua peal* 'the vase is on the table'). The multivariate results, however, indicate that once we take the variables length, complexity, and mobility of the Landmark phrase, the type of verb, and the word class of Trajector into account, the relative position between Trajector and Landmark becomes significant: when the Landmark phrase precedes the Trajector (e.g. *laual on vaas* 'on the table is a vase'), the adessive is slightly more frequently used. This result is not confirmed neither by the picture rating task nor the sentence rating task. In the picture rating task the adessive sentences where the Landmark followed the Trajector received significantly higher ratings than the corresponding *peal*-sentences. The sentence rating task demonstrates that in both word order positions, it is the adpositional construction that is preferred. As already pointed out, in the sentence rating task the variable 'length' may conflate with the semantic variable 'type of Landmark' – the adpositional construction is preferred with pieces of furniture. Naturally, the difference in results between the two tasks may have to do with the different experimental setup. The picture rating task included relatively short sentences, while the sentence rating task looked at longer and more complex environments. The precise design of corpus and experimental studies is therefore crucial – it has an effect on the results and how different studies can be compared (cf. Arppe and Järvikivi 2007).

A final point that needs to be addressed in discussing the convergence and divergence between corps and experimental results is frequency. In general, both corpus studies show that the adessive construction is many times more frequent in present-day written Estonian than the adpositional construction. Parallels can be drawn with Zipf's (1935) pioneering studies on language frequency. Zipf hypothesised that if the marked feature contains something which is absent from the unmarked, it is considered more complex and by Zipf's *principle of least effort* the more complex is used less frequently than the less complex (Greenberg 1966: 14). It is possible to argue that the adpositional construction does contain something that the case construction does not contain

– an extra word – and therefore, it is to be expected that the less complex and more concise adessive case is more frequent than the more complex and explicit adposition *peal* 'on'. It is interesting to compare the experimental results to corpus results from this perspective. Given that, overall, the adessive is much more frequent than *peal* 'on', it would be natural to assume that the former receives higher ratings than the latter in the two experimental tasks. Both rating tasks, however, show different results. In the picture rating task, the two constructions receive equally high ratings when averaged over all observations and conditions. In the sentence rating task, it is in fact the adpositional construction that receives significantly higher ratings than the adessive construction. Such results point towards the conclusion voiced by other researchers who have looked into the issue of corpus frequencies vs. acceptability judgements – care should be taken in interpreting how frequency (production) maps onto acceptability (comprehension) (cf. for example the studies reported in Kempen and Harbusch 2005; Arppe and Järvikivi 2007; Divjak 2008, in progress; Schmid 2010; Bermel and Knittl in press).

In general, employing different techniques – corpus and experimental studies – provides sufficient evidence that the morphosyntactic variables length and relative position of Trajector and Landmark do play a role in the alternation between the two Estonian locative constructions. However, the precise direction and strength of this association requires future research. It is only thanks to methodological pluralism that this dissertation found important results that lead to the re-formulation of the hypotheses and reconsidering the theoretical implications. It is specifically the diverging evidence between the results of different studies that allows the researcher to go back to the drawing board. Converging evidence enabled me to highlight some of the general trends in using the adessive and *peal*-construction in Estonian, but it is the diverging evidence that allows me to study this phenomenon further, to reconsider some of the conclusions, and to test the validity of a new set of hypotheses.

6.2 Comparison of results with other studies on grammatical synonymy

This dissertation is a first large-scale quantitative study that looks at an alternation phenomenon which is typologically different from the English word order alternations and employs a combination of both corpus and experimental methodologies with an emphasis on advanced statistical analysis. The use of "number crunching" is not taken as an end in itself, but it is seen as a vital means to achieve the ultimate goal – to provide an empirically and theoretically valid account of the alternation phenomena which has both predictive and explanatory power. The results of the dissertation demonstrate that the phenomenon I am looking at is much more difficult to capture than the English word order alternations described in Chapter 2. The studies on the English genitive alternation (Dąbrowska 1998, Gries 2002, Rosenbach 2003, Hinrichs

and Szmrecsanyi 2007, Szmrecsanyi and Hinrichs 2008, Szmrecsanyi 2010, Szmrecsanyi in press b, Shih et al. to appear), the dative alternation (Bresnan et al. 2007, Bresnan and Ford 2010, Wolk et al. to appear), particle placement (by Gries 1999, 2001, 2002, 2003a, 2003b), and comparative constructions (Mondorf 2003) are all word order alternations and it is to be expected that the differences between the alternative pairs in these alternations are more pronounced than the differences between synthetic case constructions and adpositional constructions.

Bearing in mind that the alternation between Estonian locative constructions is different from the English word order alternations, some general comparisons can still be made. The studies on the English word order alternations show that animacy and length are among the most significant explanatory variables. The multivariate corpus study reported in the dissertation does confirm the relevance of the morphosyntactic variable of length, although animacy was not a significant variable. The present studies also confirm the conclusion drawn by Hinrichs and Szmrecsanyi (2007: 464) that semantic variables are important for alternations. For the English genitive alternation, animacy is the key semantic variable (Dąbrowska 1998, Gries 2002, Rosenbach 2003, Hinrichs and Szmrecsanyi 2007, Szmrecsanyi and Hinrichs 2008, Szmrecsanyi 2010, Szmrecsanyi in press b, Shih et al. to appear); for the Estonian locative alternation, it is type and mobility of Landmark. Although Hinrichs and Szmrecsanyi (2007: 464) suggest a higher ranking for the semantic variables than for the morphological variables related to processing and parsing in their hierarchy of relevant explanatory variables, the jury is still out on the relevant ranking of different types of variables for the Estonian locative constructions, since the corpus studies and experimental studies demonstrate different results. The multivariate corpus study indicates the supremacy of length and complexity over mobility and type of Landmark, while the picture rating task and the sentence rating task point towards the conclusion that semantics is more important. A possible conclusion is that in production, language users choose the principle of language economy over other considerations, while in comprehension they, first and foremost, are guided by the semantics (i.e. type of Landmark). At the same time, the multivariate corpus analysis shows that both sets of variables (morphosyntactic and semantic) jointly do an important job in predicting the choice between the two constructions.

The results pertaining to the morphosyntactic variables studied in the dissertation point towards the conclusion that the principle of language economy plays an important role. The results of the multivariate corpus study demonstrate that cognitively more complex environments tend to favour the synthetic case construction and cognitively less complex environments the analytic adpositional construction. These results go against Rohdenburg's complexity principle (Rohdenburg 2003) and Mondorf's analytic support (Mondorf 2003) discussed in section 2.3.2. Both principles state that in case of more or less explicit constructional alternatives, the more explicit option is preferred in cognitively more complex environments. Although these principles

are stated in a very general manner and are assumed to hold cross-linguistically, it seems they may work well for the English word order alternations, but not from the perspective of the Estonian locative alternation. This result demonstrates the necessity and importance of including typologically diverse languages and different phenomena in the discussion of alternations.

The studies reported in the dissertation show converging evidence with other studies on similar alternations in Finno-Ugric languages. In reporting the univariate results of the first corpus study, reference is made to Kesküla's (2011) corpus study on the Estonian allative and the corresponding adposition *peale* 'onto' – an alternation which is similar to the adessive and adposition *peal* 'on', but instead of static location, it concerns the lative members. Both Kesküla's (2011) multivariate corpus study and the multivariate corpus study reported in this dissertation confirm the relevance of the semantic and morphosyntactic aspects of the Landmarks and the irrelevance of Trajectors to the alternation between synthetic cases and the corresponding adpositions in Estonian. The corpus and experimental results reported in this dissertation also confirm the general results found for the alternation between interior locative cases and the corresponding adpositions in the Saami languages (Bartens 1978) and the Finnish language (Ojutkangas 2008). Larger, static places prefer the case construction, while smaller, manipulable things prefer the adpositional construction. This result is in turn taken as evidence to support the more general claims made in literature that cases are more abstract than adpositions and that the former express more frequent spatial relations, while the latter less predictable spatial relations (e.g. Comrie 1986, Luraghi 1991: 66–67, Hagège 2010: 37–38, and Lestrade 2010).

6.3 Assessment and implementation of the methodology for future research on grammatical synonymy

The dissertation demonstrated the need to employ different types of methodology in studying an instance of grammatical synonymy of the kind exhibited by the synthetic case constructions and the analytic adpositional constructions in Estonian. While the corpus results may point towards one conclusion, experimental results may provide interesting contradictions that in turn lead to a better understanding of the research phenomenon on the whole. The majority of the English word order alternation studies reviewed in Chapter 2 only employ one set of methodology – corpus analysis – and although strong predictive models are fitted to the data, the question remains whether such results can be experimentally validated. Gries (2002), Rosenbach (2003, 2005), Bresnan and Ford (2010) report converging evidence, but it is in many cases the conflicting results that lead to more interesting results and reconsideration of the theoretical aspects. The field of linguistics in general needs to change its stand on publishing only results that provide converging evidence between different studies. Discussions of the results that do not validate specific hypotheses are

just as important, or even more important, than the results that provide converging evidence. Unfortunately, it is mainly the papers which report converging evidence that get published.

The general conclusion to be drawn about the implementation of the different methodological tools in the dissertation is that all of them proved to be useful with their own set of advantages and disadvantages (see Chapters 1 and 3 for overviews and discussion). Specifically, the multivariate corpus analysis is a valuable tool for studying grammatical synonymy. Such a corpus analysis allows looking at the different explanatory variables both in isolation by using various univariate statistical techniques like chi-squared test and t-test, and in combination with each other by employing binary logistic regression. Regression analysis provides information as to whether the explanatory variables that are significant in isolation, are also significant when factored in jointly. A further advantage of regression analysis is that in addition to having explanatory capacity, it also has predictive power and it provides information about the relative strength of each individual variable. The results of the regression modelling for the multivariate corpus study reported here demonstrated how the morphosyntactic variable of word order which was not significant in isolation, does play a role once the variables are factored in together.

As to the evaluation of distinctive collexeme analysis, it seems indeed useful for studying alternations, but as so many other linguistic methods, its contribution in isolation is relatively small compared to its contribution in combination with other methods like the multivariate corpus analysis and the experimental studies. In relation to studying the alternation between Estonian locative cases and adpositions, distinctive collexeme analysis along with other collostructional analysis methods may not be very cost-effective. The effort put into extracting case constructions from an untagged corpus is a feat. In addition to the relative difficulty of employing collostructional analysis for studying alternation phenomena in synthetic languages that do not have a large scale tagged corpus available, there are also more general problems associated with this set of methods. Although Gries and Stefanowitsch (2004b: 120) believe that the question whether distinctive collexeme analysis is overly sensitive in picking out words as significantly distinctive for a given construction that are not actually associated with that construction strongly enough to make predictions about their distribution can be answered in the negative, the results of the present analysis do raise some questions. For example, lexemes which are very rare in the corpus in general, come out as distinctive for the less frequent adpositional construction. Distinctive collexeme analysis and the other types of collostructional analysis is not a very good typological tool – it may be restricted to languages similar to English. Further studies are needed in assessing if and how this type of methodology can be developed to be more beneficial in studying typologically different languages and alternations which are not English word order alternations.

The choice of experimental designs was guided by the aim to include different sources of data to study the alternation between the Estonian adessive

and the adposition *peal* 'on'. Rating tasks were chosen instead of other types of experimental designs because experimental judgement data is different in nature from corpus data. In providing judgements and rating pictures or sentences, language users engage in language comprehension, while in producing either spoken or written text, they are engaged in language production. Arppe and Järvikiv (2007: 148) have pointed out that the results of the forced choice task are similar to the corpus results. According to the authors, this is to be expected, since the nature of the two datasets is similar. Acceptability judgements or other similar experimental designs (e.g. the picture rating task and the sentence rating task reported in the dissertation) may provide more useful linguistic evidence than a forced choice task or a production task. The results reported here clearly demonstrate this conclusion – methodological pluralism enabled to collect both converging as well as diverging evidence. At the same time, the results of the experimental studies described in the dissertation stress the importance of paying careful attention to how the experiments are designed – critical design issues have an effect on the results. It would be interesting to study how changing some of the aspects of the designs affects the results (e.g. changing the mode of the task – instead of the computer, to use a task on paper; changing the specific instructions – to ask the participants to judge the acceptability of the sentence rather than asking them to judge the probability or likelihood of using it themselves).

In future research on Estonian locative cases and adpositions, a larger variety of experimental studies may prove to be useful. One specific study already undertaken is to use an oral production task using the same set of stimuli and methodology as used by Lemmens and his colleagues (Lemmens and Perrez 2010) to study how the different and alternating locative constructions are used in describing different spatial relations. This technique is also used in looking at whether and how the verbal predicate affects the choice between the synthetic case construction and the analytic adpositional construction. A further possible line of enquiry is to compare the predictive power of the corpus to that of language users. A forced choice task similar to the tasks reported in Bresnan et al. (2007), Bresnan and Ford (2010) is planned, where the exact corpus sentences and contexts are given for language users with the locative construction taken out. The prediction is that the results are more or less similar, but that the model has a better overall predictive power than mirrored in the results of individual language users. Another specific experimental study already undertaken involves a more detailed study on how frequency maps onto acceptability. An acceptability judgement task is in preparation that compares the acceptability ratings for the lexemes that are identified as both distinctive and non-distinctive in the distinctive collexeme analysis reported in the thesis. The prediction is that for language users, both types of lexemes are equally acceptable and that there is no one-to-one correspondence between frequency and acceptability as demonstrated by Arppe and Järvikivi (2007), Divjak (2008, in progress), and Schmid (2010).

Another set of future research concerns corpus linguistic work. Different types of corpora are available for the Estonian language and it would be interesting to see if and how register affects the choice between the two locative constructions. Further corpus linguistic research should look at other variables besides the semantic and morphosyntactic ones reported here – phonology (e.g. stress patterns, prosody at both word and phrase level) and discourse (topicality, thematicity, lexical density, structural parallelism, persistence). In addition to looking at synchronic data, a diachronic study similar to Szmrecsanyi (in press a, in press b) may provide useful insight how the alternation has changed over time and whether Estonian is becoming more analytic and less synthetic. Yet another important aspect is lectal variation – studying the alternation between synthetic and analytic constructions on the level of the individual and taking into account the different Estonian dialects. One should also not forget the importance of polysemy when studying synonymy. A corpus-linguistic analysis of the two constructions is in progress which employs two separate collostructional analyses. These results are then compared to the results of the distinctive collexeme analysis.

6.4 Conclusion

The dissertation looks at the alternation between the synthetic case construction (the adessive case) and the analytic adpositional construction (the adposition *peal* 'on') in present-day written Estonian. Both constructions can be used to express spatial support-relations in Estonian. The thesis proceeds from the theoretical premises of cognitive linguistics and assumes that grammar and grammatical units are meaningful in their own right. The studies reported here include both corpus analysis (the multivariate corpus study and the distinctive collexeme analysis) and experimental research (two rating tasks). The focus is on the combination of different methods and looking at how the results of different sources of data lead to converging or diverging evidence. The dissertation has two aims: first, to provide a list of possible semantic and morphosyntactic explanatory variables that determine the choice between the two Estonian locative constructions, and second to assess the usefulness of the specific corpus and experimental designs as well as the statistical techniques employed. The dissertation provides a discussion on how the results obtained for a language that is typologically different from English relate to results obtained for the word order alternation that have found a relatively large-scale treatment in the literature, e.g. the genitive and dative alternation, and particle placement. The conclusion drawn is that the alternation between Estonian locative constructions exhibits a more complex phenomenon which is more difficult to capture than the English word order alternations.

It is clear that the dissertation benefited from all the three different linguistic methods described in Chapter 1 – intuition and introspection, corpus linguistics, and experimental linguistics. My own intuition as a native speaker was the key

component in deciding, first of all, which variables to examine in the dissertation, since no previous large-scale empirical study exists on the alternation between the Estonian adessive case and the adposition *peal* 'on' or any other related phenomena. Intuition and introspection were also important in interpreting the results and positing the hypotheses for the experimental tasks. Corpus linguistic analysis also served various purposes. The univariate analysis of the results enabled me to see if any of the explanatory variables I had selected was significantly associated with the construction type. The multivariate results made it possible to verify if the same set of explanatory variables is still significant once we consider the variables simultaneously and which of the variables are the strongest predictors. The corpus results, moreover, provided ample grounds for positing specific hypotheses the validity of which was tested with the two rating tasks. The experimental designs allowed me to compare how production (corpus frequencies) maps onto comprehension (language users' judgements). The combination of these different methodologies provided me with both converging and diverging evidence. This convergence and divergence of evidence reflects the complicated nature of the linguistic phenomenon under study – in comparison to the English word order alternations (e.g. particle placement, genitive and dative alternation), the typologically different alternation between the Estonian synthetic and analytic locative constructions is not as clear-cut. The linguistic evidence offered in the dissertation also confirm the conclusion that to expect full-scale convergence of evidence between different linguistic methodologies is not realistic – it is the diverging evidence that leads to important theoretical reconsiderations and the general advancement in the field.

BIBLIOGRAPHY

Agresti, Alan. 2002. *Categorical Data Analysis*. 2nd ed. New York: Wiley-Interscience.

Agresti, Alan and Barbara Finlay. 2009. *Statistical Methods for the Social Sciences*. 4th ed. New Jersey: Pearson Education International.

Ameka, Felix K. and Stephen C. Levinson. 2007. Introduction. The typology and semantics of locative predicates: posturals, positionals and other beasts. *Linguistics*, 45: 5–6, 847–872.

Andersson, John M. 1971. *The Grammar of Case: Towards a Localistic Theory*. Cambridge: Cambridge University Press.

Andersson, John M. 2006. *Modern Grammars of Case.* Oxford: Oxford University Press.

Anttila, Arto, Matthew Adams and Michael Speriosu. 2010. The role of prosody in the English dative alternation. *Language and Cognitive Processes*, 25: 7–9, 946–981.

Arnold, Jennifer E., Thomas Wasow, Anthony Losongco and Ryan Ginstrom. 2000. Heaviness vs. newness: the effects of complexity and information structure on constituent ordering. *Language*, 76: 1, 28–55.

Arppe, Antti. 2008. Univariate, bivariate and multivariate methods in corpus-based lexicography – a study of synonymy. PhD Dissertation Publications of the Department of General Linguistics, University of Helsinki,

Arppe, Antti, Gilquin Gaëtanelle, Dylan Glynn, Martin Hilpert and Arne Zeschel. 2010. Cognitive corpus linguistics: five points of debate on current theory and methodology. *Corpora*, 5: 1, 1–27.

Arppe, Antti and Juhani Järvikivi. 2007. Every method counts: combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory*, 3: 2, 131–159.

Baayen, R. Harald. 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Baayen, R. Harald. 2011. Corpus linguistics and naive discriminative learning. *Brazilian Journal of Applied Linguistics*, 11: 2, 295–328.Baayen, R. Harald, Doug J. Davidson and Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.

Bakeman, Roger. 2005. Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, 37: 3, 379–384.

Barlow, Michael and Suzanne Kemmer (eds). 2000. *Usage Based Models of Language*. Stanford, California: CSLI Publications.

Bartens, Raija. 1978. *Synteettiset ja analyyttiset rakenteet lapin paikanilmauksissa* [Synthetic and Analytic Constructions in Saami Locative Expressions]. Suomalais-ugrilaisen Seuran toimituksia 166. Helsinki: Suomalais-Ugrilainen Seura.

Bermel, Neil and Luděk Knittl. In press. Corpus frequency and acceptability judgments: A study of morphosyntactic variants in Czech. *Corpus Linguistics and Linguistic Theory*.

Bock, Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology*, 18: 3, 355–387.

Bock, Kathryn, Helga Loebell and Randal Morey. 1992. From conceptual roles to structural relations: bridging the syntactic cleft. *Psychological Review*, 99: 1, 150–171.

Bolinger, Dwight. 1968. *Aspects of Language*. New York: Harcourt.

Bolinger, Dwight. 1977. *Meaning and Form*. London: Longman.
Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Sam Featherston and Wolfgang Sternefeld (eds). *Roots: Linguistics in Search of Its Evidential Base*, 77–96. Berlin: Mouton de Gruyter.
Bresnan, Joan, Anna Cueni, Tatiana Nikitina and R. Harald Baayen. 2007. Predicting the Dative Alternation. In Gerlof Bouma, Irene Krämer, and Joost Zwarts (eds). *Cognitive Foundations of Interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.
Bresnan, Joan and Marilyn Ford. 2010. Predicting syntax: processing dative constructions in American and Australian varieties of English. *Language*, 86: 1, 186–213.
Bybee, Joan. 1985. *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: John Benjamins.
Bybee, Joan and Sandra A. Thompson. 1997. Three frequency effects in syntax. *Proceedings of the Twenty-third Annual Meeting of the Berkeley Linguistics Society*, 378–388. Berkeley: University of California Press.
Carlson, Laura A. and Patrick L. Hill. 2007. Experimental methods for studying language and space. In Monica Gonzalez-Marquez, Irene Mittelberg, Seana Coulson, and Micheal J. Spivey (eds). *Methods in Cognitive Linguistics*, 250–276. Amsterdam: John Benjamins.
Carlson, Laura A. and Gordon D. Logan. 2001. Using spatial relations to select an object. *Memory and Cognition*, 29: 6, 883–892.
Carslon-Radvansky, Laura A. and Zhihua Tang. 2000. Functional influences on orienting a reference frame. *Memory and Cognition*, 28: 5, 812–820.
Carlson, Laura A. and Emile Van der Zee. 2005. *Functional Features in Language and Space. Insights from Perception, Categorization, and Development*. Oxford: Oxford University Press.
Chafe, Wallace L. 1994. *Discourse, Consciousness, and Time*. Chicago and London: The University of Chicago Press.
Chomsky, Noam. 1957. *Syntactic Structures.* The Hague: Mouton de Gruyter.
Chomsky, Noam. 1982. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.
Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin and Use*. New York: Prager.
Chomsky, Noam. 1988. *Language and Problems of Knowledge*. Cambridge and London: MIT Press.
Christensen, Rune H. B. 2011. *A Tutorial on fitting Cumulative Link Mixed Models with the ordinal Package*. Available at http://cran.r-project.org/web/packages/ordinal/vignettes/clmm2_tutorial.pdf, accessed May 3, 2012.
Cobb, George W. 1998. *Introduction to Design and Analysis of Experiments*. New York: Springer.
Cohen, Jacob. 1992. A power primer. *Psychological Bulletin*, 112: 1, 155–159.
Cohen, Jacob. 1994. The earth is round ($p < .05$). *American Psychologist*, 49: 12, 997–1003.
Cohen, Jacob. 2009. *Statistical power analysis for the behavioral sciences*. London: Psychology Press.
Cohen, Stanley Simon. 1988. *Practical Statistics*. London: Edward Arnold.
Comrie, Bernard. 1986. Markedness, grammar, people, and the world. In Fred R. Eckman, Edith A. Moravcsik and Jessica R. Wirth (eds). *Markedness*, 85–106. New York: Plenum.

Cooper, William E. and John R. Ross. 1975. World order. In Robin E. Grossman, L. James San and Timothy J. Vance (eds). *Papers from the Parasession on Functionalism,* 63–111. Chicago: Chicago Linguistic Society.

Coventry, Kenny R. and Simon C. Garrod. 2004. *Saying, Seeing, and Acting: The Psychological Semantics of Spatial Prepositions*. New York: Psychology Press.

Coventry, Kenny R., Mercè Prat-Sala and Lynn Richards. 2001. The interplay between geometry and function in the comprehension of *over, under, above* and *below*. *Journal of Memory and Language*, 44: 3, 376–398.

Cowart, Wayne. 1997. *Experimental Syntax. Applying Objective Methods to Sentence Judgements*. London: SAGE Publications.

Crawley, Michael J. 2005. *Statistics: An Introduction Using R.* Chichester: John Wiley & Sons.

Crawley, Michael J. 2007. *The R Book*. Chichester: John Wiley & Sons.

Croft, William. 1998. Linguistic evidence and mental representations. *Cognitive Linguistics*, 9: 2, 151–173.

Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Croft, William and Alan Cruse. 2004. *Cognitive Linguistics.* Cambridge: Cambridge University Press.

Cruse, Alan. 2004. *Meaning in Language. An Introduction to Semantics and Pragmatics*. 2nd ed. Oxford: Oxford University Press.

Cuyckens, Hubert, Dominiek Sandra and Sally Rice. 1997. Towards an empirical lexical semantics. In Birgit Smeija and Meike Tasch (eds). *Human Contact Through Language and Linguistics*, 35–54. Bern: Peter Lang.

Dalgaard, Peter. 2008. *Introductory Statistics with R.* 2nd ed. London: Springer.

Dąbrowska, Ewa. 1997. *Cognitive Semantics and the Polish Dative*. Berlin and New York: Mouton de Gruyter.

Dąbrowska, Ewa. 1998. How metaphor affects grammatical coding: the Saxon genitive in computer manuals. *English Language and Linguistics*, 2: 1, 121–127.

Dąbrowska, Ewa. 2008. Questions with long distance dependencies: a usage-based perspective. *Cognitive Linguistics,* 19: 3, 391–425.

Dąbrowska, Ewa. 2010. Naive v. expert intuitions: an empirical study of acceptability judgements. *The Linguistic Review*, 27: 1, 1–23.

Divjak, Dagmar. 2006. Ways of intending: delineating and structuring near-synonyms. In Stefan Th. Gries and Anatol Stefanowitsch (eds). *Corpora in cognitive linguistics. Corpus-based Approaches to Syntax and Lexis*, 19–56. Berlin: Mouton de Gruyter.

Divjak, Dagmar. 2008. On (in)frequency and (un)acceptability. In Barbara Lewandowska-Tomaszczyk (ed). *Corpus Linguistics, Computer Tools and Applications – State of the art*, 213–233. Frankfurt: Peter Lang.

Divjak, Dagmar. 2010. *Structuring the Lexicon: a Clustered Model for Near-Synonymy.* Berlin and New York: Mouton de Gruyter.

Divjak, Dagmar. In progress. Too rare to care? Acceptability for low frequency structures is affected by logged conditional probabilities. (working title)

Divjak, Dagmar and Stefan Th. Gries. 2006. Ways of trying in Russian. Clustering behavioral profiles. *Journal of Corpus Linguistics and Linguistic Theory*, 2: 1, 23–60.

Dobson, Annette J. and Adrian G. Barnett. 2008. *An Introduction to Generalized Linear Models*. Boca Raton: CRC Press.

Dryer, Matthew S. 2005. *The World Atlas of Language Structures Online.* Available at http://wals.info/index, accessed May 25, 2008.

Dryer, Matthew S. 2007. Word order. In Timothy Shopen (ed). *Language Typology and Syntactic Description.* Vol. I. 2nd ed. Cambridge: Cambridge University Press.

Ellis, Nick C. 2002. Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24: 2, 143–188.

Erelt, Mati. 2003. Syntax. In Mati Erelt (ed). *Estonian Language*, 93–129. Linguistica Uralica. Supplementary Series. Volume 1. Tallinn: Estonian Academy Publishers.

Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael and Silvi Vare. 1993. *Eesti keele grammatika II. Süntaks* [The Grammar of Estonian II. Syntax]. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.

Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael and Silvi Vare. 1995. *Eesti keele grammatika I. Morfoloogia* [The Grammar of Estonian I. Morphology]. Tallinn: Eesti Teaduste Akadeemia Eesti Keele Instituut.

Erelt, Mati, Tiiu Erelt and Kristiina Ross. 2007. *Eesti keele käsiraamat* [Handbook of Estonian]. Tallinn: Eesti Keele Sihtasutus.

Everitt, Brian. 2005. *An R and S-PLUS® Companion to Multivariate Analysis*. London: Springer.

Everitt, Brian and Torsten Hothorn. 2010. *A Handbook of Statistical Analyses Using R*. 2nd ed. London and New York: Taylor & Francis Group.

Faraway, Julian J. 2002. *Practical Regression and Anova using R*. Available at cran.r-project.org/doc/contrib/Faraway-PRA.pdf, accessed April 28, 2012.

Featherston Sam. 2005a. The Decathlon Model of empirical syntax. In Reis Marga and Stephan Kepser (eds). *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*, 187–208. Berlin: Mouton de Gruyter.

Featherston, Sam. 2005b. Magnitude estimation and what it can do for your syntax: some WH-constraints in German. *Lingua*, 115: 11, 1525–1550.

Feist, Michele and Gentner, Dedre. 2003. Factors involved in the use of in and on. In Richard Alterman and David Kirsh (eds). *Proceedings of the Twenty-fifth Annual Meeting of the Cognitive Science Society*, 390–395. Mahwah: Lawrence Erlbaum Associates.

Fillmore, Charles J. 1982. Frame semantics. In *Linguistics in the morning calm*, 111–137. The Linguistic Society of Korea. Seoul: Hanshin.

Geeraerts, Dirk. 2006. Methodology in Cognitive Linguistics. In Gitte Kristiansen, Michel Achard, René Dirven and Francisco J. Ruiz de Mendoza Ibáñez (eds). *Cognitive Linguistics: Current Applications and Future Perspectives*, 21–49. Berlin and New York: Mouton de Gruyter.

Geeraerts, Dirk, Stefan Grondelaers and Peter Bakema. 1994. *The Structure of Lexical Variation. Meaning, Naming, and Context*. Berlin: Mouton de Gruyter.

Geeraerts, Dirk, Gitte Kristiansen and Yves Peirsman (eds). 2010. *Advances in Cognitive Sociolinguistics*. Berlin and New York: Mouton de Gruyter.

Gibbs, Raymond W. 1994. *The Poetics of Mind. Figurative Thought, Language, and Understanding.* Cambridge: Cambrdige University Press.

Gilquin, Gaëtanelle and Stefan Th. Gries. 2009. Corpora and experimental methods: a state-of-the-art review. *Corpus Linguistics and Linguistic Theory*, 5: 1, 1–26.

Givón, Talmy. 1993. *English Grammar. A Function-based Approach*. Vols. 1 and 2. Amsterdam and Philadelphia: John Benjamins.

Glynn, Dylan. 2010. Synonymy, lexical fields, and grammatical constructions. Developing usage-based methodology for Cognitive Semantics. In Hans-Jörg

Schmid and Susanne Handl (eds). *Cognitive Foundations of Linguistic Usage Patterns*, 89–118. Berlin: Mouton de Gruyter.
Glynn, Dylan and Kerstin Fischer (eds). 2010. *Quantitative Cognitive Semantics. Corpus-Driven Approaches*. Berlin: Mouton de Gruyter.
Glynn, Dylan and Justyna Robinson (eds). To appear. *Polysemy and Synonymy. Corpus Methods and Applications in Cognitive Semantics*. Amsterdam: John Benjamins.
Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Structure.* Chicago and London: University of Chicago Press.
Goldberg, Adele. 2002. Surface generalizations: An alternative to alternations. *Cognitive Linguistics*, 13: 4, 327–356.
Goldberg, Adele. 2006. *Constructions at Work: The Nature of Generalization in Language.* Oxford: Oxford University Press.
Gonzalez-Marquez, Monica, Irene Mittelberg, Seana Coulson and Michael J. Spivey (eds). 2007. *Methods in Cognitive Linguistics*. Amsterdam: John Benjamins.
Greenberg, Joseph. 1966. *Language Universals, with Special Reference to Feature Hierarchies*. The Hague: Mouton de Gruyter.
Gries, Stefan Th. 1999. Particle movement: a cognitive and functional approach. *Cognitive Linguistics*, 10: 2, 105–145.
Gries, Stefan Th. 2001. A multifactorial analysis of syntactic variation: particle movement revisited. *Journal of Quantitative Linguistics*, 8: 1, 33–50.
Gries, Stefan Th. 2002. Evidence in linguistics: three approaches to genitives in English. In Ruth M. Brend, William J. Sullivan, and Arle R. Lommel (eds). *LACUS Forum XXVIII: What Constitutes Evidence in Linguistics?,* 17–31. Fullerton: LACUS.
Gries, Stefan Th. 2003a. Grammatical variation in English: a question of 'structure vs. function'?. In Günter Rohdenburg and Britta Mondorf (eds). *Determinants of Grammatical Variation in English*, 155–173. Berlin and New York: Mouton de Gruyter.
Gries, Stefan Th. 2003b. *Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement.* London and New York: Continuum Press.
Gries, Stefan Th. 2005. Syntactic priming: a corpus-based approach. *Journal of Psycholinguistic Research*, 34: 4, 365–399.
Gries, Stefan Th. 2007a. Coll.analysis 3.2. A program for R for Windows 2.
Gries, Stefan Th. 2007b. New perspectives on old alternations. In Jonathan E. Cihlar, Amy L. Franklin and David W. Kaiser (eds). *Papers from the 39th Regional Meeting of the Chicago Linguistics Society: Vol. II. The Panels*, 274–292. Chicago: Chicago Linguistics Society.
Gries, Stefan Th. 2009a. *Quantitative Corpus Linguistics with R: A Practical Introduction.* London and New York: Routledge, Taylor & Francis Group.
Gries, Stefan Th. 2009b. *Statistics for Linguistics with R: A Practical Introduction.* Berlin and New York: Mouton de Gruyter.
Gries, Stefan Th. 2010. Behavioral profiles: a fine-grained and quantitative approach in corpus-based lexical semantics. *The Mental Lexicon*, 5: 3, 323–346.
Gries, Stefan Th. and Dagmar S. Divjak. 2009. Behavioral profiles: a corpus-based approach towards cognitive semantic analysis. In Vyvyan Evans and Stephanie S. Pourcel (eds). *New directions in cognitive linguistics*, 57–75. Amsterdam and Philadelphia: John Benjamins.
Gries, Stefan Th. and Anatol Stefanowitsch. 2004a. Co-varying collexemes in the into-causative. In Michel Achard and Suzanne Kemmer (eds). *Language, Culture, and Mind*, 225–236. Stanford: CSLI Publications.

Gries, Stefan Th. and Anatol Stefanowitsch. 2004b. Extending collostructional analysis: a corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics*, 9: 1, 97–129.
Gries, Stefan Th. and Anatol Stefanowitsch (eds). 2006. *Corpora in Cognitive Linguistics: Corpus-Based Approaches to Syntax and Lexis.* Berlin and New York: Mouton de Gruyter.
Grünthal, Riho. 2003. *Finnic Adpositions and Cases in Change*. Helsinki: The Finno-Ugrian Society.
Habicht, Külli. 2000. Grammaticalization of adpositions in Old Literary Estonian. In Mati Erelt (ed). *Estonian: Typological Studies IV*, 19–58. Tartu: Publications of the Department of Estonian of the University of Tartu 14.
Habicht, Külli. 2001a. *Eesti vanema kirjakeele leksikaalsest ja morfosüntaktilisest arengust ning Heinrich Stahli keele eripärast selle taustal* [About the Lexical and Morphosyntactic Development of Older Written Estonian and the Background for the Peculiar Features in the Language of Heinrich Stahl]. Dissertationes Filologiae Estonicae Universitatis Tartuensis 10. Tartu: Tartu Ülikooli Kirjastus
Habicht, Külli. 2001b. On the genesis and loss of the adposition *rinnas* (‘abreast, beside’) in Literary Estonian. In Ilona Tragel (ed). *Papers in Estonian Cognitive Linguistics*, 71–89. Publications of the Department of General Linguistics 2. Tartu: Tartu Ülikooli Kirjastus.
Habicht, Külli, Heiki-Jaan Kaalep, Kadri Muischnek, Kaili Müürisep and Andriela Rääbis. 2000. Kas tegelik tekst allub eesti keele morfoloogilistele kirjeldustele? Eesti kirjakeele testkorpuse morfosüntaktilise märgendamise kogemusest [Does the actual text correspond to Estonian morphological descriptions? On the experiences of the morphological tagging of the written test corpus of Estonian]. *Keel ja Kirjandus*, 9, 623–633.
Haiman, John. 1983. Iconic and economic motivation. *Language*, 59: 4, 781–819.
Haiman, John (ed). 1985. *Iconicity in Syntax*. Amsterdam: John Benjamins.
Hagège, Claude. 2010. *Adpositions.* Oxford: Oxford University Press.
Hawkins, John A. 1994. *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.
Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
Herskovits, Annette. 1986. *Language and Spatial Cognition: An Interdisciplinary Study of the Prepositions in English.* Cambridge: Cambridge University Press.
Hilpert, Martin. 2006. Distinctive collexeme analysis and diachrony. *Corpus Linguistics and Linguistic Theory*, 2: 2, 243–256.
Hinrichs, Lars and Benedikt Szmrecsanyi. 2007. Recent changes in the function and frequency of standard English genitive constructions: a multivariate analysis of tagged corpora. *English Language and Linguistics*, 11: 3, 437–474.
Jaakola, Minna. 2004. *Suomen genetiivi* [The Finnish Genitive]. Helsinki: Suomen Kirjallisuuden Seura.
Janda, Laura. 1993. *A Geography of Case Semantics: The Czech Dative and the Russian Instrumental*. Berlin: Mouton de Gruyter.
Janda, Laura. 2002. *The Case Book for Russian.* Columbus: Slavica.
Janda, Laura. 2006. *The Case Book for Czech.* Bloomington: Slavica.
Johnson, Keith. 2008. *Quantitative Methods in Linguistics*. Oxford: Blackwell.
Johnson, Mark. 1987. *The Body in the Mind. The Bodily Basis of Meaning, Imagination, and Reason*. Chicago and London: The University of Chicago Press.

Kaalep, Heiki-Jaan and Tarmo Vaino. 2000. Teksti täielik morfoloogiline analüüs lingvisti töövahendite komplektis [The full morphological analysis of a text in the linguists' toolkit], 87–99. Arvutuslingvistikalt inimesele. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised. Tartu: Tartu Ülikooli Kirjastus.

Kay, Paul and Charles J. Fillmore. 1999. Grammatical constructions and linguistic generalizations: the *What's X doing Y?* construction. *Language*, 75: 1, 1–33.

Keller, Frank. 2000. Gradience in Grammar: Experimental and Computational Aspects of Degrees of Grammaticality. Edinburgh: University of Edinburgh dissertation. Available at http://homepages.inf.ed.ac.uk/keller/publications/phd.pdf, accessed April 28, 2012.

Kemmer, Suzanne and Michael Barlow. 2000. Introduction: a usage-based conception of language. In Michael Barlow and Suzanne Kemmer (eds). *Usage Based Models of Language*, vii-xxviii. Stanford: CSLI Publications.

Kempen, Gerard and Karin Harbusch. 2005. The relationship between grammaticality ratings and corpus frequencies: a case study into word order variability in the midfield of German clauses. In Stephan Kepser and Marga Reis (eds). *Linguistic Evidence. Empirical, Theoretical, and Computational Perspectives*, 329–349. Berlin and New York: Mouton de Gruyter.

Kepser, Stephan and Marga Reis (eds). 2005. *Linguistic Evidence. Empirical, Theoretical, and Computational Perspectives.* Berlin and New York: Mouton de Gruyter.

Kesküla, Kaisa. 2011. Postpositsiooni *peale* ja allatiivi paralleelse kasutuse korpusanalüüs. [A Corpus Analysis of the Estonian Postposition *peale* 'onto' and the Allative Case Alternation]. Unpublished MA thesis. University of Tartu.

Klavan, Jane. 2008. *Medial Region Adpositions in English and Estonian: A Cognitive Grammar Perspective*. Unpublished MA thesis. University of Tartu.

Klavan, Jane. To appear. A multifactorial corpus analysis of grammatical synonymy: The case of the Estonian adessive and the adposition peal 'on'. In Dylan Glynn, Justyna Robinson (eds.). *Polysemy and Synonymy. Corpus Methods in Cognitive Linguistics.* Human Cognitive Processing. Amsterdam: John Benjamins.

Klavan, Jane, Kaisa Kesküla and Laura Ojava. 2011. Synonymy in grammar: the Estonian adessive case and the adposition *peal* 'on'. In Seppo Kittilä, Katja Västi and Jussi Ylikoski (eds). *Studies on Case, Animacy and Semantic Roles*, 113–134. Amsterdam: John Benjamins.

Laaksonen, Heli. 2000. Adessiivi suomessa ja virossa – funktiot, frekvenssit, yhtäläisyydet, erot, vastineet [The Adessive in Finnish and Estonian]. Unpublished BA thesis. Turku University.

Labov, William. 1972. Some principles of linguistic methodology. *Language in Society*, 1: 1, 97–120.

Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: The University of Chicago Press.

Lambrecht, Knud. 1994. *Information Structure and Sentence Form. Topic, Focus and the Mental Representation of Discourse Referents*. Cambridge: Cambridge University Press.

Landau, Barbara and Ray Jackendoff. 1993. "What" and "where" in spatial language and spatial cognition. *Behavioural and Brain Sciences,* 16: 2, 217–265.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Volume I: Theoretical Prerequisites*. Stanford: Stanford University Press.

Langacker, Ronald W. 2005. Integration, grammaticization, and constructional meaning. In Mirjam Fried and Hans C. Boas (eds). *Grammatical Constructions: Back to the Roots*, 157–189. Amsterdam and Philadpelphia: John Benjamins.

Langacker, Ronald W. 2007. *Ten Lectures on Cognitive Grammar by Ronald Langaker*. Gao Yuan and Li Fuyin (eds). Beijing: Foreign Language Teaching and Research Press.

Langacker, Ronald W. 2008. *Cognitive Grammar. A Basic Introduction*. Oxford: Oxford University Press.

Langemets, Margit, Mai Tiits, Tiia Valdre, Leidi Veskis, Ülle Viks and Piret Voll. 2009. *Eesti kirjakeele seletussõnaraamat.* 4 P–R [Dictionary of Written Estonian. Vol 4 P–R]. Tallinn: Eesti Keele Sihtasutus.

Lehmann, Christian. 2002. *Thoughts on Grammaticalization.* 2nd ed. Available at http://www.uni-erfurt.de/sprachwissenschaft/ASSidUE/ASSidUE09.pdf, accessed May 5, 2008.

Lehrer, Adrienne. 1982. *Wine and Conversation*. Bloomington: Indiana University Press.

Lemmens, Maarten. 2002. The semantic network of Dutch *zitten, staan,* and *liggen*. In John Newman (ed). *The Linguistics of Sitting, Standing, and Lying*, 103–139. Amsterdam and Philadelphia: John Benjamins.

Lemmens, Maarten. 2005. Aspectual posture verb constructions in Dutch. *Journal of Germanic Linguistics*, 17: 3, 183–217.

Lemmens, Maarten. 2006. Caused posture: experiential patterns emerging from corpus research. In Anatol Stefanowitsch and Stefan Th. Gries (eds). *Corpora in Cognitive Linguistics. Corpus-Based Approaches to Syntax and Lexis*, 261–296. Berlin: Mouton de Gruyter.

Lemmens, Maarten and Julien Perrez. 2010. On the use of posture verbs by French-speaking learners of Dutch: a corpus-based study. *Cognitive Linguistics*, 21: 2, 315–347.

Lestrade, Sander. 2010. *The Space of Case.* PhD dissertation, Radboud University Nijmegen.

Levin, Beth. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.

Levin, Beth and Malka Rappaport Hovav. 1995. *Unaccusativity. A Preliminary Investigation*. Chicago: Chicago University Press.

Lindström, Liina. 2005. Finiitverbi asend lauses. Sõnajärg ja seda mõjutavad tegurid suulises eesti keeles [The position of the finite verb in a clause: word order and the factors affecting it in Spoken Estonian]. Dissertationes Filologiae Estonicae Universitatis Tartuensis 16. Tartu: Tartu Ülikooli Kirjastus.

Lindström, Liina and Ilona Tragel. 2007. Eesti keele impersonaali ja seisundipassiivi vahekorrast adessiivargumendi kasutamise põhjal [Estonian Impersonal and Stative Passive, and the Use of the Adessive Argument]. *Keel ja Kirjandus*, 7, 532–553.

Lindström, Liina and Ilona Tragel. 2010. The possessive perfect construction in Estonian. *Folia Linguistica*, 44: 2, 371–400.

Lyons, John. 1977. *Semantics: 2*. Cambridge: Cambridge University Press.

McCullagh, Peter and John A. Nelder. 1989. *Generalized Linear Models*. 2nd ed. London: Chapman & Hall.

McDonald, Janet L., Kathryn Bock and Michael H. Kelly. 1993. Word and world order: semantic, phonological and metrical determinants of serial position. *Cognitive Psychology*, 25: 2, 188–230.

Maclin, M. Kimberly and Robert L. Solso. 2007. *Experimental Psychology*. 8th ed. Boston: Pearson.
Matsumura, Kazuto 1994. Is the Estonian adessive really a local case? *Journal of Asian and African Studies*, 46/47, 223–235.
Matsumura, Kazuto 1996. The Estonian adessive as 'dative'. In Heikki Leskinen, Sándor Maticsák, Tõnu Seilenthal (eds). *Congressus octavus internationalis fenno-ugristarum, Jyväskylä 10.–15.8.1995. Pars IV: Sessiones sectionum. Syntaxis et semantica & Contactus linguistici et status hodiernus linguarum & Cetera Linguistica*, 82–86. Jyväskylä: Gummerus Kirjapaino Oy.
Matsumura, Kazuto 1997. The dative use of the adessive case in Estonian: a corpus-based study. In Kazuto Matsumura and Tooru Hayasi (eds). *The Dative and Related Phenomena*, 31–79. Tokyo: Hituzi Syobo.
Mondorf, Britta. 2003. Support for *more*-support. In Günter Rohdenburg and Britta Mondorf (eds). *Determinants of Grammatical Variation in English*, 251–304. Berlin: Mouton de Gruyter.
Myers, James. 2009. The design and analysis of small-scale syntactic judgement experiments. *Lingua*, 119: 3, 425–444.
Newman, John (ed). 2002. *The Linguistics of Sitting, Standing, and Lying*. Amsterdam and Philadelphia: John Benjamins.
Newman, John and Sally Rice. 2004. English SIT, STAND, and LIE: patterns of usage and their experiential motivations. *Cognitive Linguistics*, 15: 3, 351–396.
Newman, John and Sally Rice. 2010. *Experimental and Empirical Approaches in the Study of Conceptual Structure, Discourse, and Language*. Stanford: CSLI Publications.
Norušis, Marija J. 2011. *IBM SPSS Statistics 19 Advanced Statistical Procedures Companion.* Boston: Addison Wesley.
Ojutkangas, Krista. 2001. *Ruumiinosannimien kieliopillistuminen suomessa ja virossa.* [The Grammaticalization of Body-Part Nouns in Finnish and Estonian]. Helsinki: Suomalaisen Kirjallisuuden Seura.
Ojutkangas, Krista. 2008. Mihin suomessa tarvitaan sisä-grammeja? [When are *sisä* grams used in Finnish?]. *Virittäjä*, 3, 382–400.
Ojutkangas, Krista and Tuomas Huumo. 2010. Mikä erottaa muodot sisällä ja sisässä? "Synonyymisten" muotojen analyysi [What is the difference between sisällä 'in' and sisässä 'in'? An analysis of two "synonymous" forms]. *Sananjalka*, 52, 19–45.
Palmeos, Paula. 1985. *Eesti keele grammatika II. Kaassõna* [The Grammar of Estonian II. Adposition]. Tartu: TRÜ trükikoda.
Pierce, Charles A., Richard A. Block and Herman Aguinis. 2004. Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educational and Psychological Measurement*, 64: 6, 916–924.
Pinheiro, José C. and Douglas M. Bates. 2000. *Mixed-effects models in S and S-PLUS*. New York: Springer.
Popper, Karl R. 1959. *The Logic of Scientific Discovery*. London: Hutchinson.
Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London and New York: Longman.
Rannat, Ruta. 1991. *Noomeni sünteetiliste ja analüütiliste vormide kasutus* [The Use of the Synthetic and Analytic Forms of the Noun]. Unpublished BA thesis. University of Tartu.
Rencher, Alvin C. 2002. *Methods of Multivariate Analysis*. 2nd ed. New York: John Wiley & Sons.

Rohdenburg, Günter. 2003. Cognitive complexity and *horror aequi* as factors determining the use of interrogative clause linkers in English. In Günter Rohdenburgand and Britta Mondorf (eds). *Determinants of Grammatical Variation in English*, 205–249. Berlin: Mouton de Gruyter.

Rosch, Eleanor. 1973. Natural categories. *Cognitive Psychology*, 4: 3, 328–350.

Rosch, Eleanor and Carolyn B. Mervis. 1975. Family resemblances: studies in the internal structure of categories. *Cognitive Psychology*, 7: 4, 573–605.

Rosch, Eleanor, Carolyn B. Mervis, Wayne D. Gray, David M. Johnson and Penny Boyes-Braem. 1976. Basic objects in natural categories. *Cognitive Psychology*, 8: 3, 382–439.

Rosenbach, Anette. 2003. Aspects of iconicity and economy in the choice between the *s*-genitive and the *of*-genitive in English. In Günter Rohdenburgand and Britta Mondorf (eds). *Determinants of Grammatical Variation in English*, 379–411. Berlin: Mouton de Gruyter.

Rosenbach, Anette. 2005. Animacy versus weight as determinants of grammatical variation in English. *Language*, 81: 3, 613–644.

Rosnow, Ralph L. and Robert Rosenthal. 1989. Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, 44: 10, 1276–1284.

Ross, Jaan, Pire Teras, Eva Liina Asu, Arvo Eek, Diana Krull, Ilse Lehiste, Einar Meister, Meelis Mihkla, Merike Parve and Mart Rannut. 2003. Foneetika [Phonetics]. *Emakeele Seltsi aastaraamat*, 48 (2002), 7–26.

Rudzka-Ostyn, Brygida. 1995. Metaphor, schema, invariance: the case of verbs of answering. In Louis Goossens, Paul Pauwels, Brygida Rudzka-Ostyn, Anne-Marie Simon-Vandenbergen and Johan Vanparys (eds). *By Word of Mouth. Metaphor, Metonymy and Linguistic Action in a Cognitive Perspective*, 205–244. Amsterdam: John Benjamins.

Rätsep, Huno. 1979. *Eesti keele ajalooline morfoloogia II.* [The Historical Morphology of Estonian II]. Tartu: TRÜ trükikoda.

Sandra, Dominiek. 1998. What linguists can and can't tell you about the human mind: a reply to Croft. *Cognitive Linguistics*, 9: 4, 361–378.

Sandra, Dominiek and Sally Rice 1995. Network analyses of prepositional meaning: mirroring whose mind – the linguist's or the language user's? *Cognitive Linguistics*, 6: 1, 89–130.

Schachter, Paul and Timothy Shopen. 2007. Parts-of-speech systems. In Timothy Shopen (ed). *Language Typology and Syntactic Description.* Vol. I. 2nd ed, 1–60. Cambridge: Cambridge University Press.

Schmid, Hans-Jörg. 1993. *Cottage, idea, start: Die Kategorisierung als Grundprinzip einer differenzierten Bedeutungsbeschreibung*. Tübingen: Niemeyer.

Schmid, Hans-Jörg. 2010. "Does frequency in text really instantiate entrenchment in the cognitive system?" In Dylan Glynn and Kerstin Fischer (eds). *Quantitative Methods in Cognitive Semantics: Corpus-driven Approaches*, 101–133 Berlin: Walter de Gruyter.

Schütze, Carson T. 1996. *The Empirical Base of Linguistics*. Chicago: The University of Chicago Press.

Schütze, Carson T. and Jon Sprouse. To appear. Judgment Data. In Devyani Sharma and Rob Podesva (eds). *Research Methods in Linguistics*. Available at http://www.socsci.uci.edu/~jsprouse/papers/Judgment%20data.pdf, accessed April 28, 2012.

Sharma, Devyani and Rob Podesva (eds). To appear. *Research Methods in Linguistics*. Cambridge: Cambridge University Press.

Shih, Stephanie, Jason Grafmiller, Richard Futrell and Joan Bresnan. To appear. Rhythm's role in genitive construction choice in spoken English. In Ralf Vogel and Ruben van de Vijver (eds). *Rhythm in Phonetics, Grammar and Cognition*. *Trends in Linguistics. Studies and Monographs*. Mouton de Gruyter. Available at http://stanford.edu/~bresnan/DGfS-Shihetal2010.pdf, accessed December 5, 2011.

Sprouse, Jon. 2011. A test of the cognitive assumptions of magnitude estimation: commutativity does not hold for acceptability judgments. *Language*, 87: 2, 274–288.

Sprouse, Jon and Diogo Almeida. (submitted). Power in acceptability judgment experiments and the reliability of data in syntax. Available at http://www.socsci.uci.edu/~jsprouse/#papers, accessed December 5, 2011.

Stefanowitsch, Anatol and Stefan Th. Gries. 2003. Collostructions: investigating the interaction between words and constructions. *International Journal of Corpus Linguistics*, 8: 2, 209–243.

Stefanowitsch, Anatol and Stefan Th. Gries. 2005. Covarying collexemes. *Corpus Linguistics and Linguistic Theory*, 1: 1, 1–43.

Svorou, Soteria. 1994. *The Grammar of Space*. Amsterdam: John Benjamins.

Svorou, Soteria. 2007. Relational constructions in cognitive linguistics. In Dirk Geeraerts and Hubert Cuyckens (eds). *Oxford Handbook of Cognitive Linguistics*, 726–752. Oxford: Oxford University Press.

Swerts, Marc and Carel van Wijk. 2005. Prosodic, lexico-syntactic and regional influences on word order in Dutch verbal endgroups. *Journal of Phonetics*, 33: 2, 243–262.

Szmrecsanyi, Benedikt. 2004. On operationalizing syntactic complexity. In Gérard Purnelle, Cédrick Fairon and Anne Dister (eds). *Le poids des mots. Proceedings of the 7th International Conference on Textual Data Statistical Analysis. Louvain-la-Neuve, March 10-12, 2004*. Vol. 2. Louvain-la-Neuve: Presses universitaires de Louvain. 1032–1039.

Szmrecsanyi, Benedikt. 2005. Language users as creatures of habit: a corpus-linguistic analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory*, 1: 1, 113–150.

Szmrecsanyi, Benedikt. 2006. *Morphosyntactic Persistence in Spoken English. A Corpus Study at the Intersection of Variationist Sociolinguistics, Psycholinguistics, and Discourse Analysis.* Trends in Linguistics. Studies and Monographs 177. Berlin and New York: Mouton de Gruyter.

Szmrecsanyi, Benedikt. 2009. Typological parameters of intralingual variability: grammatical analyticity versus syntheticity in varieties of English. *Language Variation and Change*, 21: 3, 319–353.

Szmrecsanyi, Benedikt. 2010. The English genitive alternation in a cognitive sociolinguistics perspective. In Dirk Geeraerts, Gitte Kristiansen and Yves Peirsman (eds). *Advances in Cognitive Sociolinguistics*. Berlin and New York: Mouton de Gruyter, 141–166.

Szmrecsanyi, Benedikt. In press a. Analyticity and syntheticity in the history of English. In Terttu Nevalainen and Elizabeth Closs Traugott (eds). *The Oxford Handbook of the History of English*. New York: Oxford University Press. Available at http://www.benszm.net/omnibuslit/Szmrecsany_ANSYNhist_Rethinking_pageproofs.pdf, accessed December 5, 2011.

Szmrecsanyi, Benedikt. In press b. The great regression: genitive variability in Late Modern English news texts. In Kersti Börjars, David Denison and Alan Scott (eds).

Morphosyntactic categories and the expression of possession. Amsterdam: John Benjamins. Available at http://www.benszm.net/omnibuslit/Szmrecsanyi_ ARCHERgenitives.pdf, accessed December 5, 2011.
Szmrecsanyi, Benedikt and Lars Hinrichs. 2008. Probabilistic determinants of genitive variation in spoken and written English: a multivariate comparison across time, space, and genres. In Terttu Nevalainen, Irma Taavitsainen, Päivi Pahta and Minna Korhonen (eds). *The Dynamics of Linguistic Variation: Corpus Evidence on English Past and Present*, 291–309. Amsterdam: John Benjamins.
Talmy, Leonard. 1983. How language structures space. In Herbert L. Pick and Linda P. Acredolo (eds). *Spatial Orientation: Theory, Research and Application*, 225–282. New York: Plenum Press.
Tannen, Deborah. 1989. *Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse*. Cambridge: Cambridge University Press.
Tauli, Valter. 1966. *Sturctural Tendencies in Uralic Languages*. London: Mouton & Co.
Taylor, John R. 2006. Polysemy and the lexicon. In Gitte Kristiansen, Michel Achard, René Dirven, Francisco J. Ruiz de Mendoza Ibáñes (eds). *Cognitive Linguistics: Current Applications and Future Perspectives*, 51–80. Berlin and New York: Mouton de Gruyter.
Teubert, Wolfgang and Ramesh Krishnamurthy. 2007. General introduction. In Wolfgang Teubert and Ramesh Krishnamurthy (eds.). *Corpus Linguistics. Critical Concepts in Linguistics. Volume I*, 1–37. London and New York: Routledge.
Tuggy, David. 1999. Linguistic evidence for polysemy in the mind: a response to William Croft and Dominiek Sandra. *Cognitive Linguistics*, 10: 4, 343–368.
Tutz, Gerhard. 2012. *Regression for Categorical Data.* Cambridge: Cambridge University Press.
Uiboaed, Kristel. 2010. Ühendverbid eesti murretes [Phrasal Verbs in Estonian Dialects]. *Keel ja Kirjandus*, 1, 307–326.
Vainik, Ene. 1995. *Eesti keele väliskohakäänete semantika kognitiivse grammatika vaatenurgast* [The Semantics of Estonian External Locative Cases from the Perspective of Cognitive Grammar]. Tallinn: Eesti Keele Instituut.
Vandeloise, Claude. 1991. *Spatial Prepositions: A Case Study from French.* Chicago: University of Chicago Press.
Veismann, Ann 2006. *peale* ja *pärast* [Estonian *peale* 'onto, on' and *pärast* 'after]. *Emakeele Seltsi aastaraamat*, 51 (2005), 170–183.
Veismann, Ann. 2008. Eesti kaassõnade tähenduste kirjeldamise probleemidest [A cognitive semantic analysis of the Estonian adpositions üle and peale: some questions and queries]. *Keel ja Kirjandus*, 5, 335–352.
Veismann, Ann. 2009. Eesti keele kaas- ja määrsõnade semantika võimalusi [Semantics of Estonian Adpositions and Adverbs]. Dissertationes Linguisticae Universitatis Tartuensis 11. Tartu: Tartu Ülikooli Kirjastus.
Wasow, Thomas. 1997. Remarks on grammatical weight. *Language Variation and Change*, 9: 1, 81–105.
Wasow, Thomas. 2002. *Postverbal Behaviour*. CSLI Publications.
Wasow, Thomas and Jennifer Arnold. 2003. Post-verbal constituent ordering in English. In Günter Rohdenburgand and Britta Mondorf (eds). *Determinants of Grammatical Variation in English*, 119–154. Berlin: Mouton de Gruyter.
Weskott, Thomas and Gisbert Fanselow. 2009. Scaling issues in the measurement of linguistic acceptability. In Sam Featherston. Susanne Winkler (eds). *The Fruits of*

Empirical Linguistics. Volume 1: Process, 229–246. Berlin and New York: Mouton de Gruyter.
Weskott, Thomas and Gisbert Fanselow. 2011. On the informativity of different measures of linguistic acceptability. *Language*, 87: 2, 249–273.
Wolk, Christoph, Joan Bresnan, Anette Rosenbach and Benedikt Szmrecsanyi. To appear. Dative and genitive variability in Late Modern English: Exploring cross-constructional variation and change. Accepted for publication in *Diachronica*. Available at http://www.benszm.net/omnibuslit/WBRS.pdf, accessed December 5, 2011.
Zipf, George K. 1935. *The Psycho-Biology of Language: An Introduction to Dynamic Philology*. Boston: Houghton Mifflin.
Zlatev, Jordan. 2007. Spatial semantics. In Hubert Cuyckens and Dirk Geeraerts (eds). *Handbook of Cognitive Linguistics*, 318 – 350. Oxford: Oxford University Press.

SUMMARY IN ESTONIAN

Korpuslingvistilised ja eksperimentaalsed meetodid grammatilise sünonüümia uurimisel

Doktoriväitekiri käsitleb eesti keele alalütlevaga ja kaassõnaga *peal* konstruktsioonide paralleelset kasutust kohasuhete väljendamisel tänapäeva kirjakeeles. Väitekiri on esimene suuremahuline kvantitatiivne uurimus grammatilisest sünonüümiast sünteetiliste ja analüütiliste vormide vahel keeles, mis on tüpoloogiliselt erinev inglise keelest. Sünonüümia uurimiseks on kasutatud töös erinevaid keeleteaduslikke meetodeid, neist olulisim on empiiriline kvantitatiivne lähenemine. Keeruliste andmeanalüüsi meetodite kasutamine ei ole olnud uurimuses eesmärk omaette, vaid vahend usaldusväärsemate tulemuste saavutamiseks. Väitekirja eemärk on anda teoreetiliselt tugev ja empiiriliselt adekvaatne ülevaade eesti keele alalütlevaga konstruktsiooni ja kaassõnaga *peal* konstruktsiooni paralleelsest kasutusest.

Kuna töö teoreetiliseks aluseks on kognitiivne keeleteadus, siis uurimus lähtub ühest selle põhiprintsiibist – et tähendus on kõikidel keeleüksustel, nii leksikaalsetel kui ka grammatilistel. Samuti lähtub töö käsitlustest, mis ei pea täielikku sünonüümiat kahe keelelise üksuse vahel võimalikuks (nt Bolinger 1968, 1977; Haiman 1983; Langacker 1987, 2008; Goldberg 2005) ning väidavad, et kui kaks üksust keeles erinevad vormiliselt, siis peavad nad erinema ka tähenduse poolest. Töö eesmärgiks on anda ülevaade nendest faktoritest, mis võiksid mõjutada adessiivi ja *peal* paralleelsest kasutust, ja välja selgitada need semantilised ja morfosüntaktilised variaablid, mis mõjutavad keelekõnelejat valima igal konkreetsel kasutusjuhul ühe sünonüümsete konstruktsioonide paarist. Kokku on töös vaadeldud 11 semantilist ja üheksat morfosüntaktilist variaablit. Töös uuriti järgnevaid semantilisi tunnuseid: trajektori ja orientiiri vahelise suhte liik, trajektori ja orientiiri elusus, arv, ja liikuvus, trajektori ja orientiiri omavaheline suhteline suurus ja kasutatud verb. Morfosüntaktilistest tunnustest käsitleti töös järgmisi: orientiirifraasi pikkus, komplekssus ja süntaktiline roll, trajektori ja orientiiri sõnaliik, trajektori kääne, lausetüüp, orientiirifraasi asend lauses, orientiiri ja trajektori omavaheline suhteline asend. Variaablite valikul said otsustavaks nii varasemate uurimuste tulemused kui uurija emakeele-intuitsioon. Vaadeldud on uurimusi, mis käsitlevad sarnast alternatsiooni teistes soome-ugri keeltes (nt Bartens 1978 ja Ojutkangas 2008 sisekohakäänete ja vastavate kaassõnade alternatsioon vastavalt saami ja soome keeles), kui ka inglise keele sõnajärjealternatsioone (nt Dąbrowska 1998, Rosenbach 2003, Hinrichs ja Szmrecsanyi 2007, Szmrecsanyi ja Hinrichs 2008, Szmrecsanyi 2010, Shih *et al*. *ilmumas* genitiivialternatsioon; Gries 1999, 2001, 2003b, Bresnan *et al*. 2007, Bresnan ja Ford 2010, Wolk *et al*. *ilmumas* daativialternatsioon).

Väitekirja tuumaks on neli uurimust: kaks korpusuuringut ja kaks katset. Katsete peamine eesmärk oli kontrollida korpusuuringute tulemuste ja intuitsiooni põhjal välja töötatud konkreetseid hüpoteese. Esimene korpusuuring on multifaktoriaalne ja selle eesmärgiks oli vaadelda kõiki 21 variaablit korraga.

Multifaktoriaalse korpusanalüüsi materjal pärineb kahest tänapäeva eesti kirjakeele korpusest. Analüüs põhineb 900 lausel: 450 alalütlevaga konstruktsiooni kasutust on võetud Tartu Ülikooli tänapäeva eesti kirjakeele korpuse morfoloogiliselt ühestatud alamkorpuse ajakirjandus- ja ilukirjandustekstidest (kokku ligikaudu 215 000 sõnet) ja 450 kaassõnaga *peal* konstruktsiooni Tartu Ülikooli tasakaalus korpuse ilukirjandustekstidest (kokku ligikaudu 5 miljonit sõnet). Esitatud on nii monofaktoriaalse kui ka multifaktoriaalse analüüsi tulemused. Monofaktoriaalse andmeanalüüsi eesmärgiks oli välja selgitada, kas ja kui tugevalt on kõik 20 variaablit seotud konstruktsiooni tüübiga (st alalütleva käändega või kaassõnaga *peal*), kui neid variaableid käsitleda iseseisvalt ja teistest variaablitest sõltumatult. Binaarne regressioon, st multifaktoriaalne andmete analüüs võimaldab aga kõigi variaablite olulisust ja mõju hinnata samaaegselt. Selline lähenemine andmete analüüsimisel peegeldab tegelikkust: reaalses keelekasutuses mõjutavad keelekasutaja valikuid kõik variaablid korraga. Regressioonanalüüs võimaldab hinnata erinevate variaablite tugevust konstruktsiooni valiku ennustamisel.

Teises korpusuuringus keskendutakse ainult semantikale ja vaadeldakse lähemalt, millist rolli mängib orientiiri liik alalütleva ja kaassõna *peal* paralleelses kasutuses. Valitud meetodiks on distinktiivne kollekseemanalüüs (*distinctive collexeme analysis*; vt nt Stefanowitsch ja Gries 2003, 2005; Gries ja Stefanowitsch 2004a, 2004b). Kollekseemanalüüs võimaldab välja selgitada, kas kahe konstruktsiooni erinevus võib olla tingitud sellest, et teatud tüüpi lekseemid eelistavad ühte kahest konstruktsioonist. Distinktiivse kollekseemanalüüsi materjal pärineb Tartu Ülikooli tasakaalus korpuse ilukirjandustekstidest (kokku ligikaudu 5 miljonit sõnet). Kokku oli selles korpuses mõlema konstruktsiooniga kasutatud 182 erinevat lekseemi. 47 lekseemi 182-st on distinktiivsed: 14 lekseemi eelistab statistiliselt oluliselt rohkem alalütlevaga konstruktsiooni ja 33 lekseemi kaassõnaga *peal*-konstruktsiooni.

Väitekirjas kirjeldatud kaks katset (piltide hinnangukatse ja lausete hinnangukatse) kontrollivad kolme hüpoteesi, mis on püstitatud korpusandmete, varasemate uurimuste ja uurija intuitsiooni põhjal. Piltide hinnangukatse vaatleb semantilist variaablit „orientiiri tüüp“ ja morfosüntaktilist variaablit „sõnajärg“. Eesmärgiks oli kontrollida kahte väidet: esiteks, et käändega konstruktsiooni eelistatakse suurte, staatiliste orientiiridega ning kaassõnaga konstruktsiooni väikeste, liigutatavate orientiiridega; ja teiseks, et adessiivi eelistatakse sellisel juhul, kui orientiirifraas eelneb trajektorifraasile ning kaassõna *peal* juhul, kui see järgneb trajektorifraasile. Kuna mõlemad konstruktsioonid väljendavad kohasuhet, siis otsustati esimeses katses kasutada visuaalset stiimulit. Töös kirjeldatud teine katse on lausete hinnangukatse ja vaatleb kahte morfosüntaktilist variaablit: orientiirifraasi pikkust ja sõnajärge. Multifaktoriaalne korpusanalüüs näitas, et orientiirifraasi pikkus on üks tugevamaid variaableid, mis mängib rolli valikul kahe konstruktsiooni vahel. Pikemate orientiirifraasidega eelistatakse kasutada sünteetilist, kohakäände konstruktsiooni. Ka teises katses uuritakse sõnajärge, kuna korpusanalüüs ja piltide hinnangukatse näitasid erinevaid tulemusi. Multifaktoriaalne korpusanalüüs annab alust arvata,

et adessiivi kasutus on veidi sagedasem juhul, kui orientiir eelneb lauses trajektorile. Piltide hinnangukatse aga näitab vastupidist. Seetõttu otsustatigi selle variaabli mõju uurida ka teises katses.

Kõik neli uurimust kinnitavad, et semantilised variaablid „orientiiri liik ja liikuvus/liigutatavus“, „orientiiri ja trajektori omavaheline suhteline suurus“ ja „verbi liik“ mängivad olulist rolli adessiivi ja kaassõna *peal* paralleelses kasutuses. Üheksast töös vaadeldud morfosüntaktilisest variaablist viis on olulised: orientiirifraasi pikkus ja kompleksus, trajektori ja orientiiri sõnaliik ja sõnajärg. Variaablid, mis käesolevas töös kirjeldatud uurimustes olulist rolli ei mängi, on trajektori erinevad semantilised omadused, trajektori kääne, orientiiri elusus ja arv, lausetüüp, trajektori ja orientiiri omavahelise suhte liik. Võib juhtuda, et kui vaadelda nende variaablite mõju koos mõne teise variaabliga, mida töös pole käsitletud, osutuvad ka need variaablid oluliseks.

Adessiivi ja kaassõna *peal* paralleelse kasutuse kohta on võimalik välja tuua järgmised üldised tendentsid:

i) **alalütlevaga konstruktsiooni** kasutatakse:
- morfoloogiliselt komplekssete ja pikkade nimisõnaliste orientiiridega (nt *kirjutuslaud*);
- staatiliste orientiiridega (nt *turg*);
- orientiiridega, mis on trajektorist suuremad.

ii) ***peal*-konstruktsiooni** kasutatakse:
- lühikeste ja lihtsate orientiirifraasidega, eriti asesõnaliste orientiiridega (nt *see*);
- väikeste ja liikuvate või liigutatavate orientiiridega (nt *kapp*);
- orientiiridega, mis on sama suured kui trajektorid;
- eksistentsiaalverbidega (nt *olema*).

Üldiselt annavad mõlemad korpusuuringud ja piltide hinnangukatse sarnaseid tulemusi. Tulemustest selgub, et orientiiri omadused (iseäranis liik ja liikuvus) mängivad olulist rolli adessiivi ja kaassõna *peal* paralleelses kasutuses. Suurte staatiliste orientiiridega eelistatakse alalütlevaga konstruktsiooni ja väiksemate, liikuvate või liigutatavate orientiiridega *peal*-konstruktsiooni. Saadud tulemused ühtivad Bartensi (1978) ja Ojutkanga (2008) uurimustega, kes leidsid sarnaseid jooni sisekohakäänete ja vastavate kaassõnade paralleelse kasutuse kohta saami ja soome keeles. Käesoleva väitekirja tulemused kinnitavad ka üldisemat tüpoloogilist väidet, et käänded on abstraktsemad ja väljendavad pigem sagedasemaid kohasuhteid (Comrie 1986, Luraghi 1991: 66–67, Hagège 2010: 37–38, ja Lestrade 2010). Samas ei ole selline eristus must-valge. Korpusuuringute tulemused näitavad, et võimalik on ka alalütleva kasutus väiksemate, liikuvate orientiiridega ja kaassõna *peal* kasutus suurte, staatiliste orientiiridega. Siiski tuleb üldine tööjaotus kahe konstruktsiooni vahel vastavalt orientiiritüübile tulemuste alusel selgelt esile ja katsete tulemused kinnitavad, et keelekasutajad on teadlikud konstruktsioonide semantilisest tööjaotusest mõisteruumis.

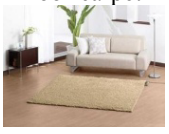
Morfosüntaktiliste variaablite osas pakuvad töös esitatud uurimused ootamatuid tulemusi. Multifaktoriaalne korpusanalüüs osutab, et orientiirifraasi pikkus on üks tugevamaid variaableid adessiiviga konstruktsiooni ja kaassõnaga *peal* konstruktsiooni paralleelses kasutuses. Sünteetilist adessiiviga konstruktsiooni eelistatakse analüütilisele kaassõnaga konstruktsioonile pikemate ja komplekssemate orientiirifraaside korral; lühikeste orientiirifraaside puhul on eelistatud kaassõnaline konstruktsioon. Motivatsioon selliseks tööjaotuseks tuleneb keeleökonoomia ja keele ikoonilisuse printsiipidest (vt nt Haiman 1983). Ühelt poolt ei soovi keelekasutajad juba niigi pikka ja keerulist fraasi veel pikemaks teha kaassõna kasutamisega ja eelistavad käändelõppu. Teiselt poolt aga on soov lühikesi ja mittespetsiifilisi fraase ikooniliselt markeerida omaette kaassõnaga, mis rõhutab kohasuhet käändelõpust paremini. Kuid lausete hinnangukatse ei kinnita korpusuuringu tulemusi. Katse tulemustest ilmneb, et katses kasutatud konkreetsete stiimulite puhul eelistasid keelekasutajad nii lühikeste kui ka pikkade orientiirifraasidega kasutada kaassõna. Neist tulemustest on võimalik teha mitmesuguseid järeldusi. Üheks järelduseks on, et katse tulemused ei peegeldanud korpusuuringu tulemusi seepärast, et stiimul sisaldas vaid ühte liiki orientiire – mööbliesemeid. Katse eesmärgiks oli hoida semantiline variaabel „orientiiri liik“ sama terves katses ning seetõttu otsustati kasutada ainult mööbliesemeid. Multifaktoriaalne korpusanalüüs ja distinktiivne kollekseemanalüüs andsid alust järeldada, et mööbliesemetega kasutatakse alalütlevaga ja kaassõnaga *peal* konstruktsiooni umbes võrdselt. Sellele vaatamata võib lausete hinnangukatse peegeldada hoopis piltide hinnangukatse tulemusi, millest selgus, et keelekasutajad eelistavad siiski mööbliesemetega kasutada pigem kaassõnaga konstruktsiooni kui kohakäänet. Seega võib oletada, et siin on semantikal suurem roll kui morfosüntaksil, ja et „orientiiri liik“ on tugevam variaabel kui „orientiirifraasi pikkus“. Edasistes uurimustes oleks huvitav vaadata, kas ja kuidas mängib orientiirifraasi pikkus rolli teist liiki trajektorite puhul.

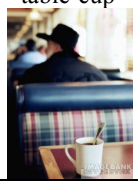
Teine osa vastupidiseks osutunud tulemusi puudutab morfosüntaktilist variaablit „trajektori ja orientiiri omavaheline suhteline asend“. Esimese korpusanalüüsi monofaktoriaalsed tulemused näitavad, et mõlema konstruktsiooniga eelistatakse sellist järjestust, kus orientiir järgneb trajektorile. See on eesti keele lause infostruktuuri seisukohast oodatav tulemus. Orientiirifraasiga väljendatakse tavaliselt diskursuse seisukohast uut infot ja see osa lausest asub eesti keeles, nagu paljudes teistes keeltes, lause lõpus. Korpusuuringu multifaktoriaalsed tulemused annavad aga alust järeldada, et kui orientiirifraas eelneb trajektorile, siis on adessiivi kasutus veidi sagedasem kui *peal* kasutus. Katsetes see tulemus aga kinnitust ei leia. Piltide hinnangukatses on neid adessiiviga lauseid, kus orientiirifraas järgneb trajektorile, hinnatud oluliselt kõrgemalt kui vastavaid kaassõnaga *peal* lauseid. Lausete hinnangukatses on mõlema sõnajärje puhul *peal*-lauseid hinnatuid oluliselt kõrgemalt kui adessiiviga lauseid. Põhjus, miks korpusuuringud ja katselised meetodid siinkohal erinevaid tulemusi annavad, võib seisneda erinevate meetodite eripäras (vrd Arppe ja Järvikivi 2007).

Viimasena esitan sagedusnäitajate tulemused. Mõlemad korpusuuringud näitavad, et adessiiviga konstruktsioon esineb tänapäeva eesti kirjakeeles lokatiivses funktsioonis märgatavalt sagedamini kui *peal*-konstruktsioon. Siinkohal võib paralleele tuua Zipfi (1935) teedrajava uurimusega sagedusest. Zipf eeldas, et kui markeeritud vormis on midagi, mis markeerimata vormis puudub, siis võib esimest pidada komplekssemaks ja Zipfi *vähima jõupingutuse printsiibi* (*principle of least effort*) järgi kasutatakse komplekssemat vormi harvemini kui vähem kompleksset vormi (Greenberg 1966: 14). Kaassõnaga konstruktsiooni võib pidada komplekssemaks konstruktsiooniks kui käändekonstruktsiooni, kuna esimeses on üks sõna rohkem – kaassõna *peal*. Seega on oodatav käände suurem sagedus võrreldes kaassõnaga. Sellest eeldusest lähtuvalt vaadati võrdlevalt korpusuuringute ja katsete tulemusi. Arvestades, et üldiselt on adessiiv sagedasem kui kaassõna *peal*, oleks loomulik eeldada, et katsetes hinnatakse adessiiviga lauseid samuti kõrgemalt kui kaassõnaga *peal* lauseid. Mõlema katse tulemused on aga sellest eeldusest erinevad. Piltide hinnangukatses on mõlemat konstruktsiooni üle kõigi eksperimentaalsete lausete ja vastuste hinnatud võrdselt kõrgelt. Lausete hinnangukatses hinnati aga just kaassõnalist konstruktsiooni oluliselt kõrgemalt kui käändekonstruktsiooni. Siinsed tulemused annavad alust nõustuda teiste uurijatega, kes on võrrelnud korpuse sagedusi vastuvõetavuse hinnangutega (nt Kempen ja Harbusch 2005, Arppe ja Järvikivi 2007, Divjak 2008, Schmid 2010, Bermel ja Knittl *ilmumas*): sagedustel põhinevate andemete tõlgendamisel tuleb olla väga tähelepanelik eriti otseste seoste loomisega ning sellega, kuidas sagedus peegeldab vastuvõetavust. Korpuse sageduste ja vastuvõetavuse hinnangute vahel ei ole üksühest vastavust.

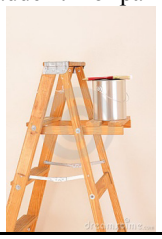
Väitekirjas esitatud uurimuste põhjal võib öelda, et mitmekesiste meetodite (katsed, korpusuuringud ja introspektsioon) rakendamine õigustas ennast. . Erinevate meetodite kasutamine andis käesolevas töös nii kooskõlalisi kui ka vastakaid tulemusi. Ühelt poolt leidis kinnitust eeldus, et eesti keele alalütlevaga konstruktsioon ja kaassõnaga *peal* konstruktsioon ei ole täissünonüümsed grammatilised konstruktsioonid ja et nende kasutuses mängivad rolli erinevad semantilised ja morfosüntaktilised variaablid. Samas andsid korpusuuringud ja katselised meetodid erinevaid tulemusi, mis lubab oletada, et kohakäänete ja -kaassõnade alternatsioon on palju keerulisem keeleline nähtus kui näiteks erinevad sõnajärjealternatsioonid inglise keeles. Viimaste puhul on kirjanduses esitatud väga selgeid tulemusi, kuid sõnajärjealternatsioonide puhul on see ka eeldatav. Töö üldisemaks panuseks keeleteaduses võib pidada seda, et vaadeldud on kahe sünonüümse konstruktsiooni vaheldust inglise keelest tüpoloogiliselt erinevas keeles. Tegemist on maailmas esimese selleteemalise kvantitatiivse uurimusega, mis rakendab nii korpusanalüüsi kui katselisi meetodeid. Metodoloogiliste tulemuste seisukohalt on olulised nii tulemused, mis erinevate meetodite teel saadud andmete põhjal ühtisid, kui ka need tulemused, mis erinesid. Just viimast liiki andmed ja tulemused on need, mis viivad uute hüpoteeside püstitamisele, seniste hüpoteeside ja eelduste muutmisele, teoreetiliste kaalutluste ümbersõnastamisele ja kriitilise mõtte arengule.

APPENDIX 1.
Picture rating task: List of stimuli

Lm-Tr pair	Sentence	Version	Experim. condition
nina-plaaster 'nose-Band-Aid' 	1. Ninal on plaaster.	version1	bodypart_lm1_ade
	2. Nina peal on plaaster.	version1	bodypart_lm1_peal
	'On the nose is a Band-Aid'		
	3. Plaaster on ninal.	version2	bodypart_lm2_ade
	4. Plaaster on nina peal.	version2	bodypart_lm2_peal
	'A/the Band-Aid is on the nose'		
õlg-kast 'shoulder-box' 	5. Õlal on kast.	version2	bodypart_lm1_ade
	6. Õla peal on kast.	version2	bodypart_lm1_peal
	'On the shoulder is a box'		
	7. Kast on õlal.	version1	bodypart_lm2_ade
	8. Kast on õla peal.	version1	bodypart_lm2_peal
	'A/the box is on the shoulder'		
põsk-lipp 'cheeck-flag' 	9. Põsel on lipp.	version1	bodypart_lm1_ade
	10. Põse peal on lipp.	version1	bodypart_lm1_peal
	'On the cheeck is a flag'		
	11. Lipp on põsel.	version2	bodypart_lm2_ade
	12. Lipp on põse peal.	version2	bodypart_lm2_peal
	'A/the flag is on the cheeck'		
selg-kivi 'back-rock' 	13. Seljal on kivi.	version2	bodypart_lm1_ade
	14. Selja peal on kivi.	version2	bodypart_lm1_peal
	'On the back is a rock'		
	15. Kivi on seljal.	version1	bodypart_lm2_ade
	16. Kivi on selja peal.	version1	bodypart_lm2_peal
	'A/the rock is on the back'		
katus-antenn 'roof-aerial' 	17. Katusel on antenn.	version1	building_lm1_ade
	18. Katuse peal on antenn.	version1	building_lm1_peal
	'On the roof is an aerial'		
	19. Antenn on katusel.	version2	building_lm2_ade
	20. Antenn on katuse peal.	version2	building_lm2_peal
	'An/the aerial is on the roof'		
põrand-vaip 'floor-carpet' 	21. Põrandal on vaip.	version2	building_lm1_ade
	22. Põranda peal on vaip.	version2	building_lm1_peal
	'On the floor is a carpet'		
	23. Vaip on põrandal.	version1	building_lm2_ade
	24. Vaip on põranda peal.	version1	building_lm2_peal
	'A/the carpet is on the floor'		

Lm-Tr pair	Sentence	Version	Experim. condition
sein-maal 'wall-painting' 	25. Seinal on maal.	version1	building_lm1_ade
	26. Seina peal on maal.	version1	building_lm1_peal
	'On the wall is a painting'		
	27. Maal on seinal.	version2	building_lm2_ade
	28. Maal on seina peal.	version2	building_lm2_peal
	'A/the painting is on the wall'		
trepp-korv 'stairs-basket' 	29. Trepil on korv.	version2	building_lm1_ade
	30. Trepi peal on korv.	version2	building_lm1_peal
	'On the stairs is a basket'		
	31. Korv on trepil.	version1	building_lm2_ade
	32. Korv on trepi peal.	version1	building_lm2_peal
	'A/the basket is on the stairs'		
laud-tass 'table-cup' 	33. Laual on tass.	version2	furniture_lm1_ade
	34. Laua peal on tass.	version2	furniture_lm1_peal
	'On the table is a cup'		
	35. Tass on laual.	version1	furniture_lm2_ade
	36. Tass on laua peal.	version1	furniture_lm2_peal
	'A/the cup is on the table'		
pink-kohver 'bench-suitcase' 	37. Pingil on kohver.	version1	furniture_lm1_ade
	38. Pingi peal on kohver.	version1	furniture_lm1_peal
	'On the bench is a suitcase'		
	39. Kohver on pingil.	version2	furniture_lm2_ade
	40. Kohver on pingi peal.	version2	furniture_lm2_peal
	'A/the suitcase is on the bench'		
tool-kott 'chair-bag' 	41. Toolil on kott.	version1	furniture_lm1_ade
	42. Tooli peal on kott.	version1	furniture_lm1_peal
	'On the chair is a bag'		
	43. Kott on toolil.	version2	furniture_lm2_ade
	44. Kott on tooli peal.	version2	furniture_lm2_peal
	'A/the bag is on the chair'		
voodi-roos 'bed-rose' 	45. Voodil on roos.	version2	furniture_lm1_ade
	46. Voodi peal on roos.	version2	furniture_lm1_peal
	'On the bed is a rose'		
	47. Roos on voodil.	version1	furniture_lm2_ade
	48. Roos on voodi peal.	version1	furniture_lm2_peal
	'A/the rose is on the bed'		

Lm-Tr pair	Sentence	Version	Experim. condition
aas-vanker 'meadow-wagon' 	49. Aasal on vanker.	version2	place_lm1_ade
	50. Aasa peal on vanker.	version2	place_lm1_peal
	'On the meadow is a wagon'		
	51. Vanker on aasal.	version1	place_lm2_ade
	52. Vanker on aasa peal.	version1	place_lm2_peal
	'A/the wagon is on the meadow'		
kallas-puu 'shore-tree' 	53. Kaldal on puu.	version1	place_lm1_ade
	54. Kalda peal on puu.	version1	place_lm1_peal
	'On the shore is a tree'		
	55. Puu on kaldal.	version2	place_lm2_ade
	56. Puu on kalda peal.	version2	place_lm2_peal
	'A/the tree is on the shore'		
põld-traktor 'field-tractor' 	57. Põllul on traktor.	version2	place_lm1_ade
	58. Põllu peal on traktor.	version2	place_lm1_peal
	'On the field is a tractor'		
	59. Traktor on põllul.	version1	place_lm2_ade
	60. Traktor on põllu peal.	version1	place_lm2_peal
	'A/the tractor is on the field'		
tänav-buss 'street-bus' 	61. Tänaval on buss.	version1	place_lm1_ade
	62. Tänava peal on buss.	version1	place_lm1_peal
	'On the street is a bus'		
	63. Buss on tänaval.	version2	place_lm2_ade
	64. Buss on tänava peal.	version2	place_lm2_peal
	'A/the bus is on the street'		
kivi-puuleht 'stone-leaf' 	65. Kivil on puuleht.	version1	thing_lm1_ade
	66. Kivi peal on puuleht.	version1	thing_lm1_peal
	'On the rock is a leaf'		
	67. Puuleht on kivil.	version2	thing_lm2_ade
	68. Puuleht on kivi peal.	version2	thing_lm2_peal
	'A/the leaf is on the rock'		
nöör-rätik 'clothesline-towel' 	69. Nööril on rätik.	version2	thing_lm1_ade
	70. Nööri peal on rätik.	version2	thing_lm1_peal
	'On the clothesline is a towel'		
	71. Rätik on nööril.	version1	thing_lm2_ade
	72. Rätik on nööri peal.	version1	thing_lm2_peal
	'A/the towel is on the clothesline'		

Lm-Tr pair	Sentence	Version	Experim. condition
raamat-kell 'book-clock' 	73. Raamatul on kell.	version2	thing_lm1_ade
	74. Raamatu peal on kell.	version2	thing_lm1_peal
	'On the book is a clock'		
	75. Kell on raamatul.	version1	thing_lm2_ade
	76. Kell on raamatu peal.	version1	thing_lm2_peal
	'A/the clock is on the book'		
redel-värvipurk 'ladder-tin of paint' 	77. Redelil on värvipurk.	version1	thing_lm1_ade
	78. Redeli peal on värvipurk.	version1	thing_lm1_peal
	'On the ladder is a tin of paint'		
	79. Värvipurk on redelil.	version2	thing_lm2_ade
	80. Värvipurk on redeli peal.	version2	thing_lm2_peal
	'A/the tin of paint is on the ladder'		

APPENDIX 2.
Sentence rating task: List of stimuli

Lm-Tr pair	Sentence	Experim. condition
diivan-padi 'couch-pillow'	1. Ilusal vanal diivanil on padi.	ade_long_lm1
	2. Ilusa vana diivani peal on padi.	peal_long_lm1
	'On the old beautiful couch is a pillow.'	
	3. Padi on ilusal vanal diivanil.	ade_long_lm2
	4. Padi on ilusa vana diivani peal.	peal_long_lm2
	'The pillow is on a/the old beautiful couch.'	
	5. Diivanil on ilus suur padi.	ade_short_lm1
	6. Diivani peal on ilus suur padi.	peal_short_lm1
	'On the couch is a beautiful big pillow.'	
	7. Ilus suur padi on diivanil.	ade_short_lm2
	8. Ilus suur padi on diivani peal.	peal_short_lm2
	'The beautiful big pillow is on a/the couch.'	
	9. See on ilus vana diivan.	control_lm
	'This is a beautiful old couch.'	
	10. See on ilus suur padi.	control_tr
	'This is a beautiful big pillow.'	
laud-vaas 'table-vase'	11. Ilusal suurel laual on vaas.	ade_long_lm1
	12. Ilusa suure laua peal on vaas.'	peal_long_lm1
	'On the big beautiful table is a vase.'	
	13. Vaas on ilusal suurel laual.	ade_long_lm2
	14. Vaas on ilusa suure laua peal.	peal_long_lm2
	'The vase is on a/the big beautiful table.'	
	15. Laual on ilus suur vaas.	ade_short_lm1
	16. Laua peal on ilus suur vaas.	peal_short_lm1
	'On the table is a/the beautiful big vase.'	
	17. Ilus suur vaas on laual.	ade_short_lm2
	18. Ilus suur vaas on laua peal.	peal_short_lm2
	'The beautiful big vase is on a/the table.'	
	19. See on ilus suur laud. 'This is a beautiful big table.'	control_lm
	20. See on ilus suur vaas. 'This is a beautiful big vase.'	control_tr

Lm-Tr pair	Sentence	Experim. condition
tool-kott ‘chair-bag’	21. Väiksel valgel toolil on kott.	ade_long_lm1
	22. Väikse valge tooli peal on kott.	peal_long_lm1
	‘On a small white chair is a bag.’	
	23. Kott on väiksel valgel toolil.	ade_long_lm2
	24. Kott on väikse valge tooli peal.	peal_long_lm2
	‘The bag is on a/the small white chair.’	
	25. Toolil on väike must kott.	ade_short_lm1
	26. Tooli peal on väike must kott.	peal_short_lm1
	‘On the chair is a small black bag.’	
	27. Väike must kott on toolil.	ade_short_lm2
	28. Väike must kott on tooli peal.	peal_short_lm2
	‘The small black bag is on a/the chair.’	
	29. See on väike valge tool. ‘This is a small white chair.’	control_lm
	30. See on väike must kott.	control_tr
	‘This is a small black bag.’	
voodi-tekk ‘bed-blanket’	31. Uuel pehmel voodil on tekk.	ade_long_lm1
	32. Uue pehme voodi peal on tekk.	peal_long_lm1
	‘On the new soft bed is a blanket.’	
	33. Tekk on uuel pehmel voodil.	ade_long_lm2
	34. Tekk on uue pehme voodi peal.	peal_long_lm2
	‘The blanket is on a/the new soft bed.’	
	35. Voodil on uus pehme tekk.	ade_short_lm1
	36. Voodi peal on uus pehme tekk.	peal_short_lm1
	‘On the bed is a new soft blanket.’	
	37. Uus pehme tekk on voodil.	ade_short_lm2
	38. Uus pehme tekk on voodi peal.	peal_short_lm2
	‘The new soft blanket is on the bed.’	
	39. See on uus pehme voodi. ‘This is a new soft bed.’	control_lm
	40. See on uus pehme tekk.	control_tr
	‘This is a new soft blanket.’	

APPENDIX 3.
The observed and expected cell counts in the two rating tasks

Table 1. Picture rating task: The observed and expected cell counts for the 20 experimental conditions and the five rating categories

Condition	Rating				
	1	2	3	4	5
ade_bodypart_lm1	37 (34.51)	28 (30.10)	20 (20.63)	20 (22.58)	41 (38.19)
ade_bodypart_lm2	22 (18.05)	20 (20.74)	15 (18.15)	22 (25.22)	67 (63.85)
ade_building_lm1	20 (16.25)	22 (19.25)	15 (17.37)	15 (24.99)	74 (68.14)
ade_building_lm2	6 (10.90)	17 (14.13)	12 (14.06)	30 (22.81)	81 (84.10)
ade_furniture_lm1	26 (26.01)	27 (26.20)	19 (20.20)	25 (24.60)	49 (48.91)
ade_furniture_lm2	13 (14.46)	19 (17.65)	18 (16.44)	23 (24.55)	73 (72.91)
ade_place_lm1	16 (15.43)	19 (18.53)	14 (16.97)	27 (24.82)	70 (70.26)
ade_place_lm2	13 (11.18)	14 (14.42)	13 (14.28)	22 (22.99)	84 (83.14)
ade_thing_lm1	49 (47.19)	28 (33.19)	21 (19.49)	23 (18.88)	25 (27.26)
ade_thing_lm2	21 (18.07)	21 (20.75)	14 (18.16)	24 (25.22)	66 (63.81)
peal_bodypart_lm1	24 (18.07)	14 (20.76)	21 (18.16)	20 (25.22)	67 (63.79)
peal_bodypart_lm2	17 (20.54)	21 (22.64)	22 (19.01)	33 (25.26)	53 (58.55)
peal_building_lm1	27 (27.53)	30 (26.99)	21 (20.37)	18 (24.31)	50 (46.80)
peal_building_lm2	27 (25.24)	20 (25.71)	24 (20.07)	26 (24.75)	49 (50.23)
peal_furniture_lm1	10 (14.34)	17 (17.54)	21 (16.38)	28 (24.51)	70 (73.22)
peal_furniture_lm2	5 (10.25)	19 (13.44)	9 (13.54)	29 (22.32)	84 (86.46)
peal_place_lm1	36 (36.27)	32 (30.72)	21 (20.57)	19 (22.08)	38 (36.36)
peal_place_lm2	36 (35.64)	30 (30.51)	19 (20.60)	25 (22.26)	36 (36.99)
peal_thing_lm1	12 (14.68)	22 (17.85)	20 (16.56)	18 (24.62)	74 (72.29)
peal_thing_lm2	8 (12.05)	15 (15.32)	17 (14.92)	29 (23.52)	77 (89.19)

Table 2. Sentence rating task: The observed and expected cell counts for the 8 experimental conditions and the seven rating categories

Condition	Rating						
	1	2	3	4	5	6	7
ade_long_lm1	72	76	73	56	47	48	40
	(66.6)	(84.7)	(67.7)	(55.8)	(52.9)	(48.0)	(36.3)
ade_long_lm2	46	72	59	60	73	59	43
	(49.3)	(70.3)	(63.4)	(58.0)	(60.7)	(60.5)	(49.7)
ade_short_lm1	21	33	22	33	58	93	152
	(14.0)	(25.4)	(31.2)	(39.5)	(60.7)	(98.4)	(142.7)
ade_short_lm2	24	54	61	55	62	85	71
	(30.9)	(49.9)	(52.2)	(55.2)	(67.3)	(79.4)	(77.0)
peal_long_lm1	31	51	56	58	71	65	80
	(32.5)	(51.9)	(53.6)	(55.8)	(66.9)	(77.6)	(73.7)
peal_long_lm2	25	40	46	50	60	94	97
	(23.4)	(39.8)	(44.6)	(50.7)	(67.4)	(88.7)	(97.3)
peal_short_lm1	7	9	23	32	55	97	189
	(8.4)	(15.9)	(20.6)	(28.2)	(48.5)	(95.6)	(194.7)
peal_short_lm2	18	37	33	45	63	101	115
	(18.9)	(33.1)	(38.7)	(46.3)	(65.6)	(94.2)	(115.3)

CURRICULUM VITAE

Name: Jane Klavan
Date of birth: 24.04.1983
Citizenship: Estonian
Telephone: +372 523 0972
E-mail: jane.klavan@gmail.com

Education:

2008–2012	University of Tartu, Faculty of Philosophy, Institute of Estonian and General Linguistics, PhD student in General Linguistics
2005–2008	University of Tartu, Faculty of Philosophy, MA (English Language and Literature)
2001–2005	University of Tartu, Faculty of Philosophy, BA (English Language and Literature)
1989–2001	Viljandi Carl Robert Jakobson Gymnasium

Professional employment:

2012	University of Tartu, Department of General Linguistics, research fellow
2011–2012	University of Tartu, Department of General Linguistics, specialist
2007–2009	University of Tartu, Department of English Language and Literature, assisstant lecturer
2005–2011	University of Tartu, Department of English Language and Literature, specialist for studies

Publikatsioonid:

Klavan, Jane. To appear. A multifactorial corpus analysis of grammatical synonymy: The case of the Estonian adessive and the adposition peal ‘on’. In Dylan Glynn, Justyna Robinson (eds.). *Polysemy and Synonymy. Corpus Methods in Cognitive Linguistics.* Human Cognitive Processing. Amsterdam: John Benjamins.

Klavan, Jane, Ann Veismann and Jürine Anni. To appear. Katselised meetodid tähenduse uurimisel [Experimental Methods for Studying Semantics]. *Journal of Estonian and Finno-Ugric Linguistics*.

Klavan, Jane, Kaisa Kesküla and Laura Ojava. 2011. Synonymy in grammar: the Estonian adessive case and the adposition *peal* ‘on’. In Seppo Kittilä, Katja Västi and Jussi Ylikoski (eds). *Studies on Case, Animacy and Semantic Roles*, 113–134. Amsterdam: John Benjamins.

Klavan, Jane, Kaisa Kesküla and Laura Ojava. 2010. Eesti keele adessiivi ja kaassõna ‘peal’ kasutus kahes keelelises katses. *Emakeele Seltsi aastaraamat*, 55, 63–86.

ELULOOKIRJELDUS

Nimi: Jane Klavan
Sünniaeg: 24.04.1983
Kodakondsus: Eesti
Telefon: 523 0972
E-mail: jane.klavan@gmail.com

Hariduskäik:

2008–2012	Tartu Ülikool, Filosoofiateaduskond, Eesti ja üldkeeleteaduse instituut, üldkeeleteaduse doktorantuur
2005–2008	Tartu Ülikool, Filosoofiateaduskond, MA (inglise keel ja kirjandus)
2001–2005	Tartu Ülikool, Filosoofiateaduskond, BA (inglise keel ja kirjandus)
1989–2001	Viljandi Carl Robert Jakobsoni nim. Gümnaasium

Töökogemus:

2012	Tartu Ülikool, Üldkeeleteaduse osakond, teadur
2011–2012	Tartu Ülikool, Üldkeeleteaduse osakond, spetsialist
2007–2009	Tartu Ülikool, Inglise filoloogia osakond, erakorraline assistent
2005–2011	Tartu Ülikool, Inglise filoloogia osakond, õppekorralduse spetsialist

Publikatsioonid:

Klavan, Jane. Ilmumas. A multifactorial corpus analysis of grammatical synonymy: The case of the Estonian adessive and the adposition peal ‘on’. In Dylan Glynn, Justyna Robinson (eds.). *Polysemy and Synonymy. Corpus Methods in Cognitive Linguistics.* Human Cognitive Processing. Amsterdam: John Benjamins.

Klavan, Jane, Ann Veismann ja Jürine Anni. Ilmumas. Katselised meetodid tähenduse uurimisel. *Eesti ja soomeugri keeleteaduse ajakiri*.

Klavan, Jane, Kaisa Kesküla ja Laura Ojava. 2011. Synonymy in grammar: the Estonian adessive case and the adposition *peal* ‘on’. In Seppo Kittilä, Katja Västi and Jussi Ylikoski (eds). *Studies on Case, Animacy and Semantic Roles*, 113–134. Amsterdam: John Benjamins.

Klavan, Jane, Kaisa Kesküla ja Laura Ojava. 2010. Eesti keele adessiivi ja kaassõna ‘peal’ kasutus kahes keelelises katses. *Emakeele Seltsi aastaraamat*, 55, 63–86.

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 p.
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 p.
7. **Larissa Degel.** Intellektuaalsfäär intellektuaalseid võimeid tähistavate sõnade kasutuse põhjal eesti ja vene keeles. Tartu, 2007, 225 p.
8. **Meelis Mihkla.** Kõne ajalise struktuuri modelleerimine eestikeelsele tekst-kõne sünteesile. Modelling the temporal stucture of speech for the Estonian text-to-speech synthesis. Tartu, 2007, 176 p.
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 määrsõnade semantika võimalusi. Tartu, 2009, 145 p.
12. **Erki Luuk.** The noun/verb and predicate/argument structures. Tartu, 2009, 99 p.
13. **Andriela Rääbis.** Eesti telefonivestluste sissejuhatus: struktuur ja suhtlus-funktsioonid. Tartu, 2009, 196 p.
14. **Liivi Hollman.** Basic color terms in Estonian Sign Language. Tartu, 2010, 144 p.