

MARIANN PROOS

Meaning and usage of Estonian experience
perception verbs



MARIANN PROOS

Meaning and usage of Estonian experience
perception verbs



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

Dissertation accepted for the commencement of the degree of Doctor of Philosophy on January 12th, 2021 by the Committee of the Institute of Estonian and General Linguistics, Faculty of Arts and Humanities, University of Tartu

Supervisors: Associate Professor Jane Klavan (University of Tartu)
Associate Professor Ilona Tragel (University of Tartu)

Opponent: Assistant Professor, PhD Laura Speed (Radboud University)

Commencement: March 26th, 2021 at 14.15 in Zoom

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



European Union
European Regional
Development Fund



Investing
in your future

ISSN 1406-5657

ISBN 978-9949-03-560-1 (print)

ISBN 978-9949-03-561-8 (pdf)

Copyright: Mariann Proos, 2021

University of Tartu Press
www.tyk.ee

ACKNOWLEDGMENTS

First of all, my deepest thanks go to my supervisors Jane Klavan and Ilona Tragel. Jane took me on as a young and excited BA student. In fact, it was a course she taught that got me so excited about science in the first place. She stuck with me through my MA studies, and took on the job of supervising me even further, during my PhD studies. She has always trusted in me as an independent researcher, even as a BA student (or at least she made me feel like an independent researcher). I am truly grateful for all the inspiration, advice, and trust she has provided. Ilona took me under her wing when I started my PhD studies and has provided me with so many various types of support during these 4 years. She was always there when I needed advice or for someone to tell me I need to re-write my article. You both have made me feel welcome to the scientific community, have made me trust in myself as a researcher and have provided invaluable advice – for this, I thank you.

Throughout my PhD years, numerous colleagues have supported me in countless ways. I was given a table in a room with Piia Taremaa and Maria Reile when I started my studies. Piia and Maria welcomed me both into the room and to a writing group, which became an invaluable resource during all my PhD years. I thank them and other members of the writing group for providing me with almost weekly feedback to my own work, and for allowing me to read and discuss theirs – this is, without a doubt, a privilege I wish every researcher had access to. I also owe thanks to my fellow PhD students with whom I got to discuss the ups and downs of research, in writing retreats and on a day-to-day basis. Special thanks go out to Mari and Anton for daily laughs, a feeling of community and continuous support: everyone should have more cake concerns in their life.

Many thanks go out to Asifa Majid and the whole Meaning, Culture and Cognition group at Radboud University in Nijmegen. The year I spent there as a part of the research group was an invaluable contribution to my growth as a researcher. I will always fondly remember my time in Nijmegen and hope I will get the opportunity to work again with the people I met during my stay, in Nijmegen or elsewhere.

Throughout my studies I have received financial support from various institutions, foundations, and projects. I thank all of them for enabling me to finish my studies: Estonian Students Fund in USA, Alfred Kordelin Foundation, Archimedes Foundation, Graduate School for Arts and Humanities, Estonian Research Council (PUT1358 “The Making and Breaking of Models: Experimentally Validating Classification Models in Linguistics”).

I also thank Mari Aigro and Jay Zameska for proofreading my various texts in English – all mistakes left are my own. Thank you, Ann for coming to my aid on the twelfth hour to edit my Estonian in the Summary. Finally, I want to

thank the kickass women I have the honour of calling my friends: Marika, Kristi, Brita, Maris, Triin, Maria, and Maiu. You are all amazing and supportive friends, mothers, professionals, and humans; thank you for being there. Last but not least, I owe thanks to my family, Aivar, Piret, and Mirjam, who have always supported me and my career choices.

To quote some of the greatest minds of this century:

*I came here to kick ass and write a dissertation,
and I'm all out of dissertation.*

In Tartu, February 4

CONTENTS

LIST OF PUBLICATIONS	8
ABBREVIATIONS	9
INTRODUCTION	10
1 BACKGROUND AND THEORETICAL STARTING POINTS	14
1.1 Cognitive semantics and polysemy	14
1.2 Cross-linguistic aspects of perception verbs and their polysemy . . .	16
1.3 Estonian perception verbs	18
2 DATA AND METHODOLOGY	26
2.1 Data and sample	26
2.2 Methodology	27
2.2.1 Sorting task	27
2.2.2 Conceptual feature rating task	28
2.2.3 Behavioural profile analysis	29
2.3 Analysis methods	31
2.3.1 Hierarchical cluster analysis (HCA)	31
2.3.2 Multidimensional scaling (MDS)	31
2.3.3 Random forest modelling	32
2.3.4 Mixed effects modelling	32
3 RESULTS AND DISCUSSION	34
3.1 Seeing the invisible (Study I)	34
3.2 Feeling your neighbour and the clock (Study II)	38
3.3 Patterns that unify and identify (Study III)	40
3.4 Discussion	43
3.4.1 Estonian experience perception verb polysemy and cross- linguistic considerations	44
3.4.2 Methodological considerations	49
4 CONCLUSION	54
SUMMARY IN ESTONIAN	57
REFERENCES	69
PUBLICATIONS	79
CURRICULUM VITAE	172
ELULOOKIRJELDUS	173

LIST OF PUBLICATIONS

Study I: Proos, Mariann (2019). Polysemy of the Estonian perception verb *nägema* ‘to see’. In: Laura J. Speed, Carolyn O’Meara, Lila San Roque, Asifa Majid (Ed.). *Perception Metaphors*, Converging Evidence in Language and Communication Research (CELCR) 19, pp. 231–252. John Benjamins Publishing Company.

Study II: Proos, Mariann (2020). Feeling your neighbour: an experimental approach to the polysemy of *tundma* ‘to feel’ in Estonian. *Language and Cognition*, 12(2): 282–309.

Study III: Proos, Mariann (2020). Usage patterns of Estonian perception verbs: A random forest approach. *SKY Journal of Linguistics*, 33: 67–101.

ABBREVIATIONS

Abbreviations in glosses

1,2,3	person	INF2	infinitive 2
APP	active past participle	NOM	nominative
CLITIC	clitic	PART	partitive
COMP	comparative	PARTICLE	particle
GEN	genitive	PL	plural
ILL	illative	PST	simple past
IMP	imperative	QUOT	quotative
INF1	infinitive 1		

Other abbreviations

BCE	Balanced Corpus of Estonian
CFR	conceptual feature rating task
etTenTen	Estonian Web Corpus
EKSS	Eesti keele seletav sõnaraamat (Explanatory Dictionary of Estonian)
HCA	hierarchical cluster analysis
MDS	multidimensional scaling

INTRODUCTION

Perception is one of the fundamental experiences of being human. Everything we experience in the world is in one way or another tied to our perceptual capabilities. This innate connection is why our language of perception is as varied as it is unified. As humans, we share the physical capability for perception, but how we conceptualise our perception experiences is influenced by our culture and our individual, unique constructions of the world. This makes perception language a fascinating topic that has been addressed by philosophers, anthropologists, and linguists alike.

This dissertation focusses on one specific means of expressing perception – perception verbs. More specifically, it considers the various meanings that these verbs have in Estonian by looking at perception verb polysemy – a phenomenon where a language item has multiple, related meanings. Since perception is so fundamental to the human experience, the experience of perception is often used to understand other, more abstract experiences. Thus, in English, we can *see a cat* but we can also *see a problem*. We conceptualise *seeing a problem* through the notion of physical experience by carrying over some characteristics of the physical experience to the abstract experience. For example, we tend to trust our sense of sight over our other senses and consider it more objective than our other senses. Similarly, *seeing a problem* demands a degree of objectivity. You can make someone see the problem by explaining it to them. In a similar way you could make someone see the same thing that you are seeing by putting them in the same physical position you are in and pointing in the direction of the object.

However, whilst this might seem logical for an English speaker, it might not be so for someone speaking a different language. The way we use perception verbs is very much shaped by the nature of our language and even our culture. For example, in Estonian, it makes perfect sense to *feel your neighbour* meaning that you know your neighbour very well – this would likely make much less sense to an English speaker than *seeing a problem*. This dissertation aims to explore the different meanings that Estonian perception verbs have, especially in the context of cross-linguistic perception verb meaning patterns. In addition, the dissertation also looks at how these verbs are used as represented by corpus material. The dissertation thereby offers a unique contribution to the understanding of how different languages use perception language to conceptualise various experiences.

Researching meaning is far from a straight-forward task. The question of how to research something that does not exist outside of the language user's mind is not an easy one – the present dissertation does not provide a conclusive answer to this. Nonetheless, semanticists strive towards providing insights as to how the language user's mind might work and making inferences about the meaning that is hidden in it, as does this dissertation. This dissertation approaches polysemy from both experimental and corpus-linguistic viewpoints and aims to map possible methodological difficulties that might arise in regards to meaning research.

The dissertation sets off from two main lines of research. The first line concerns the polysemy of five Estonian perception verbs: *nägema* ‘to see’, *kuulma* ‘to hear’, *tundma* ‘to feel’, *maitset tundma* ‘to taste’ and *lõhna tundma* ‘to smell’ in a cross-linguistic context. The second line concerns the methods of researching polysemy.

The research questions the dissertation answers are as follows:

1. what are the various meanings that Estonian experience perception verbs carry? (Q1);
2. how is the polysemy of experience perception verbs organised in the language user’s mind? (Q2);
3. how does Estonian experience perception verb polysemy relate to cross-linguistic findings about experience perception verbs and their polysemy? (Q3);
4. which usage patterns are characteristic to Estonian experience perception verbs? (Q4);
5. which considerations need to be addressed in regards to the methods of polysemy research? (Q5).

Q1. The first research question is answered throughout Studies I–III. A combination of experimental (Studies I and II) and corpus-based methods (Studies I and III) is used together with insights from Estonian dictionaries. The goal of Q1 is not to provide an exhaustive list of meanings each experience verb has. Rather, this dissertation aims to describe and analyse the types of meanings the perception verbs carry. The results regarding Q1 are relevant to the study of Estonian and offer valuable insights into its nature.

Q2. In addition to finding out which meanings Estonian experience perception verbs carry, this dissertation also aims to investigate how the polysemy of perception verbs is organised in the language user’s mind. In Studies I and II this is established through behavioural experiments that offer insights as to how meaning is organised in the native speaker’s mind. In Study III, a corpus-based method is employed to the same end. The methods used in this dissertation are not claimed to provide direct access to the human mind. However, by gathering evidence with multiple methods we can piece together a reflection which, step by step, leads us closer to how meaning is organised in the mind.

Q3. The third research question aims to place Estonian experience perception verbs and their polysemy into a larger cross-linguistic context. This question is also answered throughout Studies I–III. By comparing findings about Estonian experience perception verb polysemy to cross-linguistic evidence we gain knowledge about the multitude of ways perception is expressed through language. This is important not only for the study of language, but also offers valuable insights

to how different cultures conceptualise the experience of perception.

Q4. The fourth research question is dealt with in Study III. Describing and analysing the usage patterns of Estonian experience perception verbs serves as a tool for both gaining knowledge about how a language works and for gaining knowledge about how the mind of the speaker works. By describing how an element is used in a language, we can come closer to answering the question of why it is used in such a manner.

Q5. The fifth and final research question concerns the methods of researching polysemy. Throughout the dissertation, various methods are applied to polysemy research. Methodologically, this dissertation is a unique contribution to meaning research. It not only applies different methods, but combines the results from these methods to put together a comprehensive picture about meaning. Experimental methods are used in Studies I and II, and corpus-based methods in Studies I and III. Meaning is an elusive subject to study and each method comes with its own advantages and disadvantages, both of which are analysed throughout the dissertation.

The present dissertation consists of two main parts: the introductory part and the publications that make up the body of the dissertation. The introductory part consists of six sections. First, the background for the study is laid out and its theoretical starting points are introduced. The dissertation follows the framework of cognitive linguistics and usage-based linguistics. These frameworks are introduced in turn in section 1.1 with special regard to the approach to polysemy in them.

Next, in section 1.2, main findings about perception verbs in other languages are introduced. In addition, an overview of the general approach to perception verb research is given. This allows to put the results concerning Estonian perception verbs into a larger cross-linguistic context. Section 1.3 is an introduction to the Estonian perception verb paradigm. The notion of “main perception verbs” is introduced here and it is shown how the verbs that the dissertation focusses on relate to other Estonian perception verbs. In addition, consideration is given to other Estonian verbs that belong to the field of perception vocabulary. A selection of corpus samples is introduced to offer an overview of the extent of polysemy among Estonian main perception verbs.

In section 2, the data that the dissertation is based on is presented, along with an overview of the methods the dissertation employs. In section 2.1, both the experimental as well as corpus-based material are introduced. This section offers an overview of the size and the composition of the data samples. Next, the methodology employed in the dissertation is introduced in sections 2.2.1, 2.2.2, and 2.2.3. Here, experimental methods (sorting task and conceptual feature rating task) and the corpus-based method are introduced separately. In section 2.3 the methods used for data analysis are explained, including hierarchical cluster

analysis (HCA), multidimensional scaling (MDS), random forest modelling and mixed effects modelling.

Results of the studies included in the dissertation are introduced in section 3. Each of the three studies (Study I, Study II, and Study III) and their results are introduced separately. In section 3.4, the results from all the studies are discussed. The discussion has two parts: the first part (section 3.4.1) discusses the results that relate to the polysemy of Estonian perception verbs (research questions Q1–Q4) and the second part (section 3.4.2) focusses on the methodological implications (research question Q5). The introductory part finishes with section 4 which reiterates the most important findings of the dissertation, places them in a wider context, and offers ideas for future research.

1 BACKGROUND AND THEORETICAL STARTING POINTS

In this section, some theoretical starting points and general background are presented. This dissertation follows a cognitive linguistic and usage-based approach to language. In Section 1.1 some main points on how polysemy is understood in cognitive semantics are introduced. The distributional approach is also introduced in Section 1.1. In Section 1.2, an overview of perception verbs in a cross-linguistic context is given, as well as an overview of patterns of perception verb polysemy. The last section offers an overview of Estonian perception verbs. Consideration is given to the paradigm of main perception verbs in Estonian, their meaning variation, and their representation in the corpus.

1.1 Cognitive semantics and polysemy

Polysemy is a language phenomenon in which a language element has multiple related meanings. This encompasses both verb-noun pairs such as ‘murder’ (v) and ‘murder’ (n), cases of typical metonymy such as ‘school’ (institution) and ‘school’ (building) as well as more complex networks like the verb ‘run’ in English, e.g. *the river runs past the house* and *he runs very fast*.

The present dissertation adheres to a cognitive linguistic approach to polysemy. At the heart of cognitive linguistic theory lies the view that language is not an independent module in the brain, but rather is fundamentally connected to our cognitive processes, our senso-perceptual and senso-motoric systems (Langacker, 1987, 1991; Talmy, 2000, 2003; Gibbs, 2006). Thus, human beings’ language capacities are inherently tied to our experiences of the world around us. Furthermore, the meaning of a linguistic unit is composed of both lexico-grammatical aspects and encyclopaedic knowledge, i.e. world-knowledge and knowledge of other concepts (Fillmore, 1985; Lakoff, 1987; Talmy, 2000).

The process of a language item acquiring multiple polysemous meanings is understood via the notion of semantic extension – a process where a language item that denotes one concept (source) is used to denote another concept (goal). For this to be possible, there needs to be some underlying conceptual similarities between the source and the goal. These conceptual similarities are based on schematic representations of our world experience – fundamental patterns of how we make sense of the world (Lakoff and Johnson, 1980; Johnson, 1987; Lakoff, 1987). Quite often, the semantic extensions we see in a language follow the direction physical field → abstract field, i.e. abstract concepts are lexicalised via linguistic means that are used to lexicalise physical, concrete concepts. For example, using ‘run’ in the sense of ‘trying to be elected’ is an example of an abstract use of ‘run’.

In cognitive semantics, polysemy is understood as categorical in nature. The

meanings that an item exhibiting polysemy has are called *senses*¹ and they form a network of categories. These categories can consist of multiple senses, some more prototypical than others, which follows the idea of *family resemblance* (Rosch and Mervis, 1975, p. 575). The boundaries between these different categories are not discrete; rather, the network is a fluid structure with frequent overlaps (Rosch, 1978; Geeraerts, 2006). As Lakoff (1987, p. 16) discusses, the idea of non-traditional categories (i.e. categories with frequent overlaps) that drives the understanding of polysemy in cognitive semantics stems from Wittgenstein's famous work (Wittgenstein, 1953). There are no discrete lexical entries in the mind of a language user, each representing one and only that sense. Rather, polysemy is characterised by a complex structure consisting of overlapping categories with non-discrete borders.

Considering the complexity of polysemy, the context in which the polysemous word is used in becomes one of the key parts of interpreting its meaning. The importance of context is tied to the principles of usage-based linguistics, which cognitive semantics abides by (see e.g. Gries, 2006; Gries and Divjak, 2009; Glynn and Fischer, 2010; Glynn and Robinson, 2014). Usage-based theories set off from the principle that all language is language in use, and every usage has a function. The linguistic context that an expression occurs in can tell us something about the function of that expression. Or in the famous words of John Rupert Firth: "you shall know a word by the company it keeps" (Firth, 1957, p. 11). The idea that the function of a linguistic expression can be derived from the contexts that it is used in, is known as the distributional hypothesis (Harris, 1954). As meaning is, in essence, a type of function, we can look at the distribution of linguistic items in order to draw inferences about their meanings (Gries and Divjak, 2010, p. 338).

Specifically, for polysemy this means that a word's meanings are its different functions. Within a polysemy network, meanings that are more similar occur in similar linguistic contexts. According to what Lenci (2008, p. 16) calls the strong distributional hypothesis, the distribution of the contexts of a word has a causal role in the formation of the semantic content of that word. The present dissertation adheres to the distributional hypothesis to the extent that context and distributional patterns are an integral part of the semantic content of an expression. However, it refrains from establishing a unidirectional causal link between the two.

A final note has to be made about cognitive semantics and the empirical commitment. Since usage is the key to understanding any linguistic function, usage-based linguistics takes a strong data-driven approach to research. Inferences about language can only be drawn on the basis of the actual use of language, and data that is sufficiently representative of the population (Barlow and Kemmer, 2000; Croft and Cruse, 2004; Tummers et al., 2005; Glynn, 2010). Moreover, the usage-

¹ In the present dissertation, I have tried to minimise the use of the term *sense* in this meaning. This is to avoid confusion with the term *sense* as in *sense of sight* which is also used throughout the dissertation. Thus, *meaning* and *polysemous meaning* are used to express this concept not as a theoretical choice, but as a pragmatic one.

based approach is strongly aligned with the principles of good empirical science. As such, claims about language should be falsifiable and the research process reproducible. The present dissertation also conforms to an empirical data-based research method by only using authentic data and making both the data and the analysis scripts freely available.

1.2 Cross-linguistic aspects of perception verbs and their polysemy

Although physical perception skills are the same for humans as a species, the way we make sense of our perception experiences varies geographically, culturally, and linguistically. The differences in sensory perception across the world have been discussed i.a. by Howes (1991, 2005) and Classen (1993, 2005). As a part of our cognitive system, language offers a unique look into this variation. A plethora of evidence suggests that the lexical means of expressing perception are as varied as the cultures themselves. Languages differ in how (or whether) they lexically differentiate between the senses as well as which types of perception they prioritise in their culture.

Viberg (1984) has proposed a system of main perception verbs that is useful to characterise the perception verb lexicon across languages (Viberg, 1984, p. 125). This paradigm contains five sense modalities: sight, hearing, touch, tasting, and smelling. The paradigm is further split into what Viberg calls dynamic systems: activity, experience, and source-based (Viberg, 1984, p. 128). The dynamic systems refer to the type of perception within one sense, e.g. whether the perceiver is an active agent or not. Viberg (1984) considers the experience vs activity distinction as the difference between a state and a controlled activity. This describes pairs like *see* vs *look*, where the first one describes a non-volitional experience, and the other an experience that has a human agent and is consciously controlled. The copulative or source-based verb differs from the previous two modes in that it places the source of the experience in the role of the subject (Viberg, 1984), e.g. ‘the picture looks funny’, where the picture is the subject of the clause.

Languages can have different strategies for “filling the slots” in the main perception verb paradigm. For example, languages differ in how many senses are separately lexicalised – there are languages that have one verb for seeing and another verb for all the other senses, e.g. Avatime, Duna, Italian, Spanish, Tzeltal, and Whitesands (San Roque et al., 2014, p. 41), and languages that lexicalise all senses separately, such as English. The dynamic systems are also lexicalised differently in different languages. English, for example, exhibits what Viberg calls “complete lexical differentiation” in the modalities of sight and hearing, which means that each dynamic system has a separate verb in these modalities. For the modalities of touch, taste, and smell, there is only one verb that covers all the dynamic systems (Viberg, 1984, p. 128).

The way languages lexicalise perception experiences has been shown to vary cross-culturally, and although vision does seem to be universally prioritised in languages (San Roque et al., 2014), other sense modalities show differential lexicalisation. As already mentioned, some languages only make the distinction between seeing and ‘other perception’. Some languages have lexical inventories for perception experiences that other languages do not. For example, most European languages do not have abstract smell terms. Instead, we describe scents with constructions like “smells like X” or borrow terms from the gustatory domain, i.e. “smells sweet”. However, it is not the case that smells are difficult to lexicalise in principle. An ever-growing body of research is now showing that olfaction and its lexicalisation is highly culture-specific, and some cultures have an extensive vocabulary related to the sense of smell, including abstract smell terms (Majid and Burenhult, 2014; Wnuk and Majid, 2014; Croijmans and Majid, 2016; Majid and Kruspe, 2018).

Perception verbs also tend to be polysemous across languages, and there are some universal tendencies as to what types of more abstract experiences are lexicalised through perception verbs in languages. Ibarretxe-Antuñano (2008) has proposed a general perceiving-is-understanding connection. Variants of this association have been found across the world’s languages in very different forms. In Indo-European languages, vision seems to be the most productive field through which concepts related to understanding and intelligence are lexicalised (Sweetser, 1990). In Australian languages, the association is stronger and more productive between audition and understanding (Evans and Wilkins, 2000). Viberg (2005, 2015) showed that in Swedish, Norwegian, and Finnish, the tactile modality is one of the main sources for the perceiving-is-understanding metaphor. Even smell vocabulary has been shown to carry cognition-related meanings (Storch, 2013). Thus, it seems that the connection between perception and cognition is incredibly pervasive across very different languages.

Similarly to the notion of understanding, social interaction has also been observed to be lexicalised through perception vocabulary. For example, audition vocabulary has been observed to also lexicalise concepts like obeying someone or following someone’s advice (Evans and Wilkins, 2000; Vanhove, 2008). Social interaction in terms of meeting someone and spending time with them, or even having a relationship with someone is often lexicalised through vision verbs (Evans and Wilkins, 2000; Storch and Aikhenvald, 2013b; San Roque et al., 2018).

Although we can observe the aforementioned shared patterns of semantic extensions across languages, there are some patterns that are more language specific (Storch and Aikhenvald, 2013a; Speed et al., 2019; O’Meara and Majid, 2020). While the motivation behind shared tendencies is generally based on shared cognitive abilities, language-specific patterns are tied more to the culture and conventions of a language. The language of perception is quite uniquely positioned at the crossroads of cognition, culture, and language. The cross-linguistic variation of

perception language not only enables us to look at different lexicalisation patterns, but also to look at different conceptualisation patterns in light of different cultures (Majid and Levinson, 2011; Majid et al., 2018).

While research on the topic of perception vocabulary in European languages is quite extensive, the focus has mostly been on Indo-European languages. Estonian holds a rather unique position in Europe, and it is thus extremely valuable as an addition to the typological knowledge we currently have. Estonian is a Finno-Ugric language with approximately one million speakers. However, historically, Estonian has been greatly influenced by Low German, Russian, and to a smaller extent, Finnish (Erelt, 2003). On the one side, thus, the Finno-Ugric heritage plays an important role in the characteristics of the language, especially on the level of grammar. On the other side, the vocabulary of Estonian is largely influenced by its contacts with Low German and Russian.

This dissertation is a step toward filling the gap that currently exists in perception verb research on European languages outside the Indo-European family. In addition, it aims to advance the knowledge about Finno-Ugric languages in general. Thirdly, this dissertation offers a valuable contribution to the research on perception verbs in general, i.e. in regards to the different ways perception is understood and expressed cross-linguistically.

1.3 Estonian perception verbs

Over the years, there has been extensive research into sense perception language in Estonian. Most of the research has focussed on Estonian colour terms (Sutrop, 2002; Uusküla, 2008; Uusküla et al., 2012; Uusküla and Sutrop, 2014; Uusküla and Bimler, 2020). Vilja Oja's work on the topic is perhaps the most influential and extensive; some recent works include Oja (2001); Oja and Uusküla (2010); Oja (2014) and Oja (2016). Color terms have even been investigated in Estonian sign language by Hollman and Sutrop (2010). Other sensory modalities have not received as much attention, some exceptions are works on smell (Sutrop, 2001, 2002) and taste (Sutrop, 1998, 2002) vocabulary. Based on these works, it is certain that the colour vocabulary of Estonian is rich and varied. However, as in so many other European languages, smell and taste vocabulary shows considerably less variation and richness.

Estonian perception verbs, however, have received minimal attention. For example, Born (1995) has looked at Estonian perception verbs in a diachronic setting. Labi (2006) offers an analysis of some Estonian perception verbs in the context of Estonian runic songs; she found, for example, that in runic songs, seeing, hearing, and feeling all function as sources for semantic extension into the field of cognition, and in the case of feeling, the cognition-related meaning is most frequent. Sepper (2006) has also discussed the use of some Estonian perception verbs in fiction and non-fiction at the end of the 19th and the beginning

of the 20th century. Sepper found that the verbs *nägema* ‘to see’, *kuulma* ‘to hear’ and *tundma* ‘to feel’ are direct evidentials in Estonians, while intransitive verbs *näima* ‘to seem’, *paistma* ‘to be visible / to seem’ and *tunduma* ‘to seem’ are indirect evidentials.

This section offers an overview of the main perception verb paradigm in Estonian. What is classified as a perception verb can vary from author to author, but the example of Viberg’s seminal work is often followed, which established a paradigm of the main perception verbs (Viberg, 1984, p. 125) (see Section 1.2). A proposed paradigm of Estonian main perception verbs is presented in Table 1, based partly on my own intuition as a native speaker, and in part on information available from dictionaries, discussed below.²

Table 1: Proposed Estonian perception verb paradigm

	Experience	Activity	Source-based
Visual	<i>nägema</i> ‘to see’	<i>vaatama</i> ‘to look’	<i>välja nägema</i> (out + see.INF2) ‘to look’
Auditory	<i>kuulma</i> ‘to hear’	<i>kuulama</i> ‘to listen’	<i>kõlama</i> ‘to sound’
Tactile	<i>tundma</i> ‘to feel’	<i>katsuma</i> ‘to touch’	<i>tunduma</i> ‘to feel’
Gustatory	<i>maitset tundma</i> (taste.PART + feel.INF2) ‘to taste’		<i>maitsuma</i> ‘to taste’
Olfactory	<i>lõhna tundma</i> (smell.PART + feel.INF2) ‘to smell’	<i>nuusutama</i> ‘to smell’	<i>lõhnama</i> ‘to smell’

As Table 1 shows, Estonian exhibits complete lexical differentiation in all three dynamic systems in the case of four out of the five sense modalities: sight, hearing, touch, and smell. Taste is lexically differentiated between experience and activity/source-based systems. The experience verbs of tasting and smelling are extended through the tactile perception verb *tundma* ‘to feel’. These are compound verbs, formed by adding the respective object (either *lõhna* ‘smell.part’ or *maitse-t* ‘taste-part’) to the verb. Notice also that the source-based verb in the sight modality is also a compound verb – *välja nägema*. This particle verb consists of the verb particle *välja* ‘outward’ and the experience verb *nägema* ‘to see’. The verb does not need to occur in an adjacent position to the noun phrase or the particle in a sentence.

Table 1 is not a depiction of all perception verbs in Estonian. The main perception verbs are verbs that are semantically neutral, i.e. they have no connotations and are not marked for lexical aspect. For example, *puudutama* ‘to touch’ is not considered the main activity perception verb in the modality of touch because it is marked for lexical aspect and refers to a momentary action. Similarly, *haisema*

² There are two infinitives in Estonian – *da*-infinitive (INF1) and *ma*-infinitive (INF2). *ma*-infinitive is sometimes called supine in Estonian grammar. Here and elsewhere I use the *ma*-infinitive form (INF2) of the verb, which is the conventional base form of Estonian verbs and also the citation form.

‘to stink’ is not a main perception verb in the source-based dynamic system for the olfactory modality, because it only refers to bad smells.

Still, there are a number of possible alternative lexical units to the ones presented in Table 1, especially in the “source-based” modality. For example, the visual modality could potentially also be expressed by verbs *näima* ‘to seem’, *paistma* ‘to look’ or the construction *näha olema* ‘be visible’. All of these lexical units can express a source-based visual experience in certain contexts, similarly to *välja nägema* ‘to look’. Which of these can be deemed objectively a main perception verb, would require separate research, however, it is possible to discuss the different semantic aspects of the verbs based on dictionary information and preliminary corpus material.

For example, EKSS lists two meanings for *välja nägema* ‘to look’: the direct physical meaning of ‘being able to look outside’³ and the more abstract ‘having an appearance’. In addition, the EKI Combined Dictionary lists the meaning ‘to look, to work in a certain way, to seem’ (e.g. *Milline näeb välja sinu tavaline tööpäev?* ‘How does a typical workday look for you?’). For *näima*, EKSS lists one contemporary meaning: ‘to look like / seem as something/someone’ (e.g. *Ta näib vanemana kui ta tegelikult on* ‘(S)he looks older than (s)he is’). *Paistma* has the primary meaning of ‘to shine/to warm up’ (e.g. *Päike paistab soojalt* ‘The sun is shining warm’), but the meaning ‘be visible (from afar)’ is also listed (e.g. *Linna tornid juba paistavad* ‘The towers of the city are already visible’). The construction *näha olema* is considered a specific type of passive construction in the Estonian Reference Grammar (Erelt and Metslang, 2017), but semantically, it can also denote a source-based visual experience (e.g. *Maja oli juba kaugelt näha* ‘The house was visible from a great distance’).⁴

In the source-based dynamic system of the auditory modality, *kostma* ‘to sound’ (*Koridoris kostsid kellegi sammud* lit. ‘In the corridor, someone’s steps were audible’ ‘You could here someone walking in the corridor’) could be considered as an alternative to *kõlama* ‘to sound’. *Kostma* ‘to sound’ and *paistma* ‘to be visible’ are parallel forms in the sense that they both carry an additional dimension of distance, i.e. they would not generally be used to denote perception experiences where the stimuli are nearby. That being said, there are certainly exceptions to this and the conventions for choosing either variant deserve further research. In conclusion, although further research is required, it seems that there are a number of verbs that can be considered to occupy the slots for the source-based dynamic systems for the visual and auditory modalities, and which is the main perception verb is not an easy question to answer.

Similarly, in the experience dynamic system of the olfactory modality, *haistma*

³ The particle *välja* ‘to the outside’ is also polysemous. It can combine with a verb to make a particle verb with a new meaning, but it can also retain its directional meaning. In the latter case, the particle denotes the direction of the activity.

⁴ See also Huumo (2010) for a discussion about a similar construction in Finnish, named ‘perceptibility verbs’ in Huumo (2010).

‘to smell’ could be considered as an alternative to *lõhna tundma* ‘to smell’. A preliminary corpus search from the etTenTen corpus showed 559 tokens with *haistma* in comparison to 1465 for *lõhna tundma*. However, we do know that frequency information alone is not enough to decide over any kind of primacy or entrenchment (e.g. Schmid, 2010), and so although the corpus frequencies suggest that *lõhna tundma* is used more, it does not confirm that this is a main perception verb. Thus, the paradigm presented in Table 1 is only one possible representation of the main perception verb paradigm, and to some extent is only reliant on my own instinct as a native speaker.

In addition to the many alternatives possible to the ones presented in Table 1, *tunduma* ‘to feel’ in the source-based dynamic system is an especially interesting case. Although it is used to express the source-based tactile experience, it is also very frequently used to encode other sensory experiences. A more appropriate translation to English would thus be ‘to seem’, encoding a general type of perception which can combine multiple sensory experiences. It also denotes experiences beyond the traditional sensory ones, but can also express ‘having a feeling or opinion about something’ (EKSS). As will be shown in Sections 3.2 and 3.4, this is tied to the root *tund(i)-* and its etymology.

Although this dissertation takes under observation only the experience perception verbs, it is useful to consider them in the context of all the other main perception verbs. To this end, the Internet corpus of Estonian (etTenTen)⁵ was used. All of the main perception verbs (n=14) were searched for in the user interface of the corpus and sentence counts were compared. This gives an overview of where Estonian perception verbs stand cross-linguistically (frequency-wise) as well as enables us to put the experience perception verbs in the context of other main perception verbs.

⁵ The corpus contains 270 million words from 686 000 websites. The domains of the websites are as follows: forum, government, religion, periodicals, blog, informative, and unknown. The domains allow for a relatively large range of different texts, including more official written language as well as language more similar to spoken language (i.e. forums, blogs).

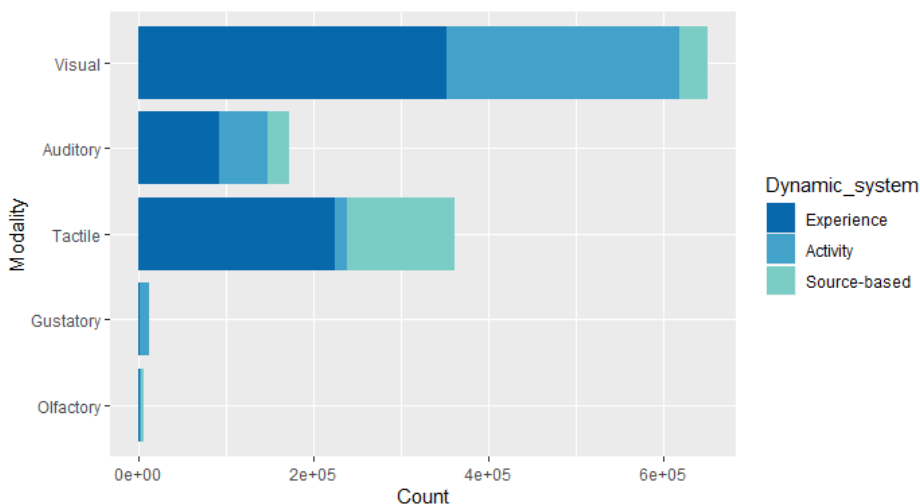


Figure 1: Absolute counts of the main Estonian perception verbs in the etTenTen corpus

Figure 1 shows that vision verbs were the most frequent, followed by tactile, auditory, gustatory, and olfactory verbs.⁶ The dominance of vision verbs coincides with the findings of San Roque et al. (2014, p. 49), who showed that vision verbs occurred the most frequently in their conversational corpus, which contained data from 13 languages around the world. Although not a universal trend, they also found the auditory modality was the second most prominent in many languages in their sample. This was not the case for the Estonian corpus data reflecting the frequencies of the main perception verbs. Instead, tactile perception verbs strongly hold the second position. The absolute frequencies were 649,968 for the visual, 361,257 for the tactile, and 172,335 for the auditory modality. This means that there were roughly two times fewer tactile verbs than vision verbs and a further two times fewer auditory verbs than tactile verbs. The counts are thus drastically different. The counts for the gustation and olfaction verbs were even smaller at 11,831 and 5,663 respectively.

All the main perception verbs have more than one meaning listed in the Explanatory Dictionary of Estonian (EKSS) (Langemets et al., 2009). The number of meanings for the verbs resembles the overall corpus frequencies to a large degree.⁷

⁶ Since 3 of the main perception verbs are compound verbs (*välja nägema* ‘to look’, *lõhna tundma* ‘to smell’, *maitset tundma* ‘to taste’) the sentence counts do not reflect the situation with 100% accuracy. Because the compound verbs are separable, i.e. the lemmas can occur in different parts of sentences, it is necessary to search for sentences where both of the lemmas are simply present (as only searching for sentences where the lemmas occur next to each other would exclude a lot of matches). However, this means that the sentence counts include sentences where these words do not form a compound verb, reducing the accuracy of the count.

⁷ The number of meanings presented here excludes compound verbs that are listed as independent entries in EKSS (e.g. *kaasa tundma* lit. ‘with feel’ ‘to sympathise’ or *pealt kuulma* lit. ‘off.of hear’ ‘to overhear’).

Verbs of visual perception have altogether 29 meanings listed in the EKSS: twelve meanings for *nägema* ‘to see’, fifteen meanings for *vaatama* ‘to look’, and two meanings for *välja nägema* ‘to look’. Verbs of tactile perception have 26 meanings listed: eighteen for *tundma* ‘to feel’, six for *katsuma* ‘to touch’, and two for *tunduma* ‘to feel’. Auditory perception verbs have considerably fewer meanings, with only 15 in total: seven for *kuulma* ‘to hear’, five for *kuulama* ‘to listen’, and three for *kõlama* ‘to sound’. For verbs of gustatory and olfactory perception, EKSS does not list the experience verbs *maitset tundma* ‘to taste’ and *lõhna tundma* ‘to smell’ as independent entries, but rather includes them under the meanings of *tundma* ‘to feel’. For the other dynamic systems, EKSS lists three meanings for *maitsuma* ‘to taste’, which covers both activity and source-based verbs in the gustatory modality. For the olfactory modality, the activity verb *nuusutama* ‘to smell’ has three meanings, and the source-based verb *lõhnama* ‘to smell’ has two meanings.

Thus, considering the main perception verbs, vision verbs in Estonian are overall the most frequent and have the most meanings as well. Vision verbs are followed by tactile verbs, which are followed by auditory verbs, both frequency-wise and by number of meanings in EKSS. The trend is reversed for verbs of gustatory and olfactory perception: while gustation verbs hold the second to last place frequency-wise, they have the smallest number of meanings. Vision thus holds its prominent position in the framework of Estonian perception verbs quite strongly: vision verbs in general are the most frequent as well as most polysemous, which confirms the primacy of vision once again (see also San Roque et al. (2014)). However, the single most polysemous main perception verb is not a vision verb, but rather the tactile verb *tundma* ‘to feel’ with 18 meanings.

To get an idea about how these differences in levels of meaning variation are reflected in language usage, a 3,000-sentence sample was extracted from the etTenTen corpus. The aim was to assess the proportion of perception meanings vs non-perception meanings of the verbs. The sample consists of 200 sentences for each modality in each dynamic system (e.g. 200 for *nägema* ‘to see’ (experience), 200 for *vaatama* ‘to look’ (activity), 200 for *välja nägema* ‘to look’ (source-based)). All of the sentences were annotated as expressing a sense-meaning (expressing physical perception) or some kind of other meaning. The coding was coarse, i.e. only cases where the physical sense-meaning was unambiguous were counted as expressing sense-meaning. The amount of perception vs non-perception meanings per sense modality is presented in Figure 2. The dynamic types on the figure are conflated, i.e. “visual” includes the verbs *nägema* ‘to see’, *vaatama* ‘to look’ and *välja nägema* ‘to look’.

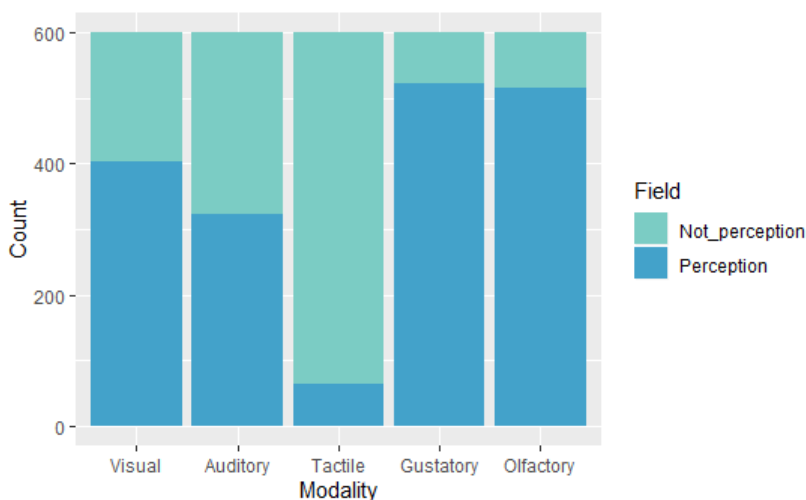


Figure 2: Counts of perception and non-perception meanings of Estonian perception verbs in a corpus sample

Figure 2 shows that there are more perception meanings in the 3,000-sentence corpus sample. However, it also shows how large the number of non-perception meanings is for some of the modalities. Although for most of the perception verbs the absolute frequency of perception meanings surpasses that of non-perception meanings, this is not the case in the tactile modality. For the verbs of the tactile modality, perception meanings make up only 11% ($n = 64$) of the total 600 sentences. In comparison, the gustatory and olfactory modalities exhibit the opposite trend: perception meanings make up 87% ($n = 523$) in the gustatory and, almost equally, 86% ($n = 516$) in the olfactory modality.

Interestingly, while vision verbs are the most frequent and have the largest number of meanings based on dictionary information and the absolute frequency data, the present 3,000 sentence sample does not reflect this: both the tactile perception verbs and auditory perception verbs exhibit more polysemous meanings in the sample. This might, again, hint to the primacy of the visual perception modality, e.g. San Roque et al. (2014) also found that vision is the most talked about sense in their conversational corpus. On the other hand, it might also reflect the fact that some meanings of tactile perception verbs are very pervasive in the sample, i.e. there might be one or even several meanings in which tactile verbs are used very often.

The data presented here effectively paints a picture of the Estonian perception verb system as being dynamic and varied. Although there is still much to be discovered about each and every perception verb in the Estonian language, the present work focusses on the experience perception verbs in Estonian: *nägema* ‘to see’, *kuulma* ‘to hear’, *tundma* ‘to feel’, *maitset tundma* ‘to taste’, and *lõhna tundma* ‘to smell’. These verbs form a coherent set in that they can be compared

to each other and analysed as conceptually expressing a similar type of experience. At the same time, the set is not too big for a comprehensive analysis that still remains within the limits of a dissertation. This being said, it would without a doubt be an interesting endeavour to look at all the main perception verbs as a full set. As was shown in the present section, the verbs in the source-based dynamic system and their alternatives would also offer fruitful research opportunities in the future. The present dissertation is the first to focus on Estonian perception verbs and thus leaves ample room for further research on the topic.

Study I takes a thorough look at the visual perception verb *nägema* ‘to see’ and its polysemy. *Nägema* ‘to see’ is the most frequent main perception verb out of all the perception modalities and dynamic systems. Together with the attested primacy of visual perception vocabulary in a cross-linguistic context (see Section 1.2), *nägema* ‘to see’ makes for an excellent case study both in the framework of Estonian perception verbs as well as in a cross-linguistic context. Study II focusses on *tundma* ‘to feel’ as the perception verb with the richest polysemy out of all the main perception verbs in Estonian. In Study III, all of the experience perception verbs are brought together in an extensive corpus study. The study looks at the polysemy of these verbs in the context of larger overall usage patterns that emerge from the corpus, exploring the interplay of form and meaning.

2 DATA AND METHODOLOGY

In the following sections, the data that the present dissertation is based on and the methodology that is used for data collection and analysis are introduced. First, an overview of both the corpus samples and experimental datasets is given in Section 2.1. This is followed by an introduction of the methodology in Section 2.2. Section 2.2 first introduces the research methods (i.e. data collection methods) and then moves on to introduce the analysis methods.

2.1 Data and sample

The data included in this dissertation is collected from experiments as well as from corpora. All of the datasets from the studies included in the dissertation are published and freely accessible through the Open Science Framework (Proos, 2019a, 2020b,a).

Table 2: Experimental samples used in the dissertation

	STUDY I	STUDY II	
	Sorting task	Sorting task	Conceptual feature rating task
No. of participants	20 (F 11, M 9)	66 (F 53, M 13)	68 (F 61, M 7)
Age of participants	mean 24 yrs. range 20–50	mean 31 yrs. range 21–65	mean 29 yrs. range 18–57
No. of exp. items	sentences: 13	sentences: 25	sentences: 72 dimensions: 6

The experimental data consists of the results of two sorting tasks and a conceptual feature rating task as represented by Table 2. The sorting task was used in Studies I and II, and the conceptual feature rating task in Study II. In Study I participants sorted 13 sentences based on the meaning of the verb *nägema* ‘to see’ in the sentence. The dataset from the second sorting task (Study II) consists of sorting data of 25 sentences containing the verb *tundma* ‘to feel’. The third experimental dataset (Study II) consists of rating data from the conceptual feature rating task. The dataset includes ratings for 72 sentences on 6 different dimensions using a 7-point Likert scale.

Table 3: Corpus samples used in the dissertation

	STUDY I	STUDY III
Corpus	Balanced corpus of Estonian (BCE)	Estonian Internet corpus (etTenTen)
Size of corpus	15 mln words	270 mln words
Size of sample	700 sentences	2418 sentences

The corpus-based material consists of two datasets, as shown in Table 3. The first dataset (Study I) is a 700-sentence random sample from the Balanced Corpus of Estonian (BCE, 2014). The corpus consists of 15 million words with equal samples from fiction, non-fiction, and scientific texts. In Study I, scientific texts were excluded from the sample. Altogether, the second corpus sample (Study III) includes 2,418 sentences. The sample was collected from the etTenTen corpus. The corpus consists of 686,000 web pages in Estonian with a total of 270 million words.

2.2 Methodology

This dissertation employs multiple research and analysis methods. One of the goals of this research was to apply both experimental and corpus-based methods to the study of polysemy. Applying multiple methods allows the researcher to get a more complete picture of the topic at hand: all methods have their trade-offs and supplementing the different methods with each other helps bridge that gap.

The methods used in this dissertation all follow a bottom-up approach to research, i.e. generalisations are made on the basis of a collection of individual observations. The bottom-up approach goes hand in hand with the principles of usage-based linguistics, as the actual instance of language usage is deemed to be the most appropriate analysis material.

In the following sections, the research methods as well as the analysis methods used in the studies which are included in the present dissertation are laid out. All of the statistical analysis for this dissertation was conducted via the statistical software R (R Core Team, 2019). All of the scripts used for the statistical analysis are published and freely accessible (Proos, 2019a, 2020b,a).

2.2.1 Sorting task

Studies I and II apply the sorting task method to the research of polysemy. The theory behind the sorting task stems from categorisation: language users tend to place together items they perceive as more similar. Thus, when presented with sentences with a polysemous item in its various meanings, participants should sort sentences where the verb has a similar meaning into one group. Since the theory

of polysemy in cognitive semantics states that similar meanings have similar underlying conceptual characteristics, the task reflects how meaning is organised in the language user's mind. The design has been used for polysemy research i.a. by Sandra and Rice (1995), Beitel et al. (2001), Gibbs and Matlock (2001), and Raukko (2003).

The goal of employing the sorting task for polysemy research in the dissertation was to determine which types of meanings language users perceive as more similar, i.e. which meaning groups form from the list of meanings. According to cognitive semantics theory, meanings with similar underlying conceptual patterns should be perceived as more similar (Johnson, 1987). Thus, by letting language users sort meanings (represented by sentences) into groups, it is possible to determine which meanings are more similar at a conceptual level.

In the sorting task, language users are given a list of language elements and their task is to categorise these elements into groups. There are numerous variations of the design: one may have a set number of groups the participants have to form; there may be a time limit to the task, or it may be possible to sort one element into multiple groups at the same time. The current dissertation uses this method in the same manner in both Studies I and II. Language users had to sort sentences into groups according to the meaning of the verb that was indicated to them in the instructions (either *nägema* 'to see' (Study I) or *tundma* 'to feel' (Study II). They could sort the sentences into as many groups as they wanted; there was no time limit, and afterwards they had to name the groups they created.

The sentences were chosen from a corpus sample: in Study I, the sample was from the Balanced Corpus of Estonian (BCE), and in Study II, it was from the etTenTen corpus. The list of the sorting task sentences was compiled by manually going through corpus examples and choosing sentences to represent each meaning of the verb. The goal was to choose sentences that were, on the one hand, as simple as possible, but at the same time, most representative of the corresponding meaning. The lists of sentences can be found in Studies I and II, included in this dissertation.

2.2.2 Conceptual feature rating task

Study II uses a conceptual feature rating (CFR) task in the research of the polysemy of *tundma* 'to feel'. The CFR task is a method developed by Troche et al. (2014, 2017) as a method for describing concepts on an abstractness-concreteness scale. The concepts are represented by language items and language users are asked to rate these items on various cognitive dimensions, represented by Likert scales. Every concept thus gets a value for each dimension, allowing for a multi-dimensional representation of the concepts.

This dissertation applies a modified version of this task. The goal of the task in Study II was to find out whether native Estonian speakers rate sentences with

tundma ‘to feel’ in its cognition-related meanings (e.g. *Jaan tunneb füüsikat hästi* ‘Jaan knows physics well’) as more concrete than sentences with other cognition verbs such as *teadma* ‘to know’ or *aru saama* ‘to understand’.

Native Estonian speakers had to rate sentences on the cognitive dimensions of time, emotion, social interaction, action, sensation, and thought. Troche et al. (2014) also included dimensions for polarity, space, quantity, morality, ease of modifying, and ease of teaching⁸. However, for the present study, a smaller amount of dimensions was included firstly because the participants had to rate sentences not single words, making the task more time consuming: inclusion of too many dimensions would have made the task too strenuous for the participants. Secondly, a subjective choice was made in regards to which type of dimensions would be the most informative considering the subject matter at hand, e.g. the dimension of morality was not included for this reason. Proos (2020c, p. 14–17) offers a more thorough overview of the dimensions included, and the resulting hypotheses.

2.2.3 Behavioural profile analysis

The behavioural profile method (Gries, 2006, 2010) can be used to reach various research goals. The rationale for the method stems from the distributional hypothesis (see Section 1.1). Shortly, by looking at how a language item behaves (i.e. in which contexts it occurs) in a corpus, we can make inferences about its function. In the present dissertation, the method is employed to fill two different research aims. Firstly, it is used to look at which polysemous meanings of the verb *nägema* ‘to see’ group together according to characteristics observable from the corpus sample, thereby making it possible to look at underlying patterns of meaning extension (Study I). Secondly, the behavioural profile method is used to look at usage patterns of the five perception verbs under study as a class of verbs, with the aim of seeing whether and to which extent it is possible to predict the verb in the sentence with the help of contextual cues present in the corpus (Study III).

The behavioural profile analysis method has been used for polysemy research i.a. by Berez and Gries (2008); Gries and Divjak (2009); Glynn (2014); Jansegers et al. (2015); Jansegers and Gries (2020). The method consists of the following steps. First, a representative corpus sample is collected. In this sample, the linguistic unit in question functions as the dependent variable. This sample is annotated for many possible different independent variables, from morphosyntactic to discourse variables. The variables are selected in accordance with the language element that is studied, i.e. there are no pre-set rules as to which variables should be included in the study. The co-occurrences of the dependent variable with every

⁸ The rationale for including the chosen dimensions in the CFR design is laid out in Troche et al. (2014: 2).

independent variable are then annotated. This process results in frequency data, which is then analysed via different statistical methods, which can be explorative (e.g. hierarchical clustering, multidimensional scaling) or confirmatory (e.g. regression analysis) in nature.

Coding schema. The corpus data in Studies I and III was annotated for various morphosyntactic and semantic variables. The morphological variables concern the perception verb in the sentence and were manually coded in Study I, and automatically coded, but checked manually in Study III. The syntactic and semantic variables were annotated in a bottom-up manner. This means that there were no *a priori* variables, rather, manually going through the data allowed for repeating patterns to emerge that were then operationalised as variables. The annotation schema for Studies I and III is shown in Table 4. As can be seen from the table, significantly more variables were used in Study I. This is due to the differential nature of the studies: while Study I was an in-depth look at the polysemy of *nāgema* ‘to see’, Study III focussed on overall patterns of five different perception verbs.

Table 4: Annotation schema of the corpus samples from Study I and Study III

VARIABLE TYPE	VARIABLES	Study I	Study III
TEXT TYPE	fiction, non-fiction	X	
SEMANTIC	meaning of verb	X	
	animacy of source	X	
	abstractness of source	X	
	abstractness of stimulus	X	X
	semantic type of stimulus	X	X
	semantic type of stimulus lemma	X	
	semantic tense	X	
	semantic person	X	
	SYNTACTIC	type of modifier	X
serial verb		X	
type of sentential complement		X	
construction type			X
MORPHOLOGICAL	person	X	X
	tense	X	X
	polarity	X	X
	mode	X	X
	voice	X	X

2.3 Analysis methods

In this section, the analysis methods used in the dissertation are introduced. The methods include both exploratory statistical methods and confirmatory statistical methods. Each method is briefly introduced. For analysing the experimental data included in this dissertation, hierarchical cluster analysis (HCA), multidimensional scaling (MDS), and mixed effects modelling are used. For analysing the corpus-based data, hierarchical cluster analysis and random forest modelling are used.

2.3.1 Hierarchical cluster analysis (HCA)

Hierarchical cluster analysis is an exploratory method for finding structure in data. The goal is to find items that are grouped together on the basis of some similarities or common characteristics in the data. Thus, from the data clusters are formed that include items that are more similar to each other than to other items in other clusters. Different parameters which concern how and how many clusters are formed may be set before running the analysis. Divjak and Fieller (2014) offer a thorough overview of the applications of HCA in cognitive linguistics.

This dissertation employs an agglomerative clustering method. The analysis starts with all the given items as separate clusters. The algorithm then adds the most similar ones together, and this process is repeated until all the items are added into a single cluster. This method is used in Study I to visualise the results of both the sorting task and the behavioural profile analysis. By applying a clustering method, the patterns that emerged from the participants sorting decisions can be analysed, which in turn are indicative of a level of mental organisation of the polysemous meanings.

2.3.2 Multidimensional scaling (MDS)

Multidimensional scaling is an exploratory technique for visualising structure in data (see Levshina (2015) for applications in linguistics). Similarly to HCA, the goal is to visualise similarities between items based on common characteristics. The visualisation is based on the distances of data points from each other. Each data point gets a coordinate on a set number of axes and is then plotted in the two- or three-dimensional space. In essence, both HCA and MDS are tools for exploring patterns and visualising similarities in data. The main difference between the two is in the categorisation of objects. With HCA, each item is assigned to a cluster⁹, i.e. an item cannot belong to two clusters on the same hierarchical level.

⁹ With HCA, items can belong to multiple clusters in a hierarchical manner: since the algorithm starts by considering each item as a separate cluster and then agglomerates them into larger ones, each item can belong to a number of higher level clusters.

MDS does not form distinct groups; rather, it offers a concentration-cloud like structure, where some of the items are closer to each other and some further apart. This makes the representation more dynamic.

Multidimensional scaling is used in Study II to visualise the results from the sorting task concerning the polysemy of *tundma* ‘to feel’. The distances between data points are calculated based on language users’ sorting decisions. Thus, the closer the data points are on the resulting map, the more the items represented by the data points were sorted together. Similarly to the hierarchical clustering method, MDS is also used to find patterns in sorting data, which are indicative of the organisation of the elements in the language user’s mind.

2.3.3 Random forest modelling

Random forest modelling (Breiman, 2001; Tagliamonte and Baayen, 2012) has been gaining a lot of popularity in linguistics because it is capable of handling “weird” data well. In corpus-based language studies, the data may often be un-balanced. For example, there may be a disproportionately small number of observations on some level of an independent variable. The so-called tree-and-forest models are well suited to analyse such un-balanced data, as they do not blow the importance of these variables out of proportion (Strobl et al., 2008). A single conditional inference tree is based on a partitioning algorithm. The algorithm repeatedly splits data into two so that observations with similar response variables are grouped together (Strobl et al., 2009). The partitioning is continued until pre-set conditions are reached. The random forest is a collection of single inference trees. By using a random sampling method and comparing observed values with the actual values, the random forest algorithm calculates the importance of each predictor, i.e. which variables are the most important in predicting the response variable (Tagliamonte and Baayen, 2012).

Random forest modelling is used in Study III. The goal of Study III was to find out, which predictors could be the most important ones in determining the meaning of perception verbs in a sentence. By applying random forest modelling, the predictors included in the study were analysed and it was shown which ones predict the verb in the sentence most successfully. This, in turn, is indicative of the different types of information that can be derived from context to determine meaning.

2.3.4 Mixed effects modelling

Mixed effects models are, in essence, regression models that allow including both fixed and random effects into one model (Winter, 2013; Gries, 2013; Baayen, 2014). Random effects are variables which do not cover all possible levels of the population. For example, the participant is a random effect, since we are dealing

with a random sample of people. Over the past years, starting as early as Clark (1973), there has been increasingly more attention drawn to the fact that language is not a fixed effect either – similarly to the participant, the experimental items we use in linguistics are merely a tiny sample of language in its entirety. Thus, it is useful to always consider the experimental items to be random effects. In the present study, mixed effects modelling is used in Study II to determine whether the choice of verb (*tundma* vs. other) had an effect on the participant's ratings in the conceptual feature rating task.

3 RESULTS AND DISCUSSION

In this section, the main results from the studies are introduced. Study I deals with the polysemy of *nägema* ‘to see’ using both experimental (a sorting task) and corpus-based (behavioural profile analysis) methods. The goal of the study was to determine how native Estonian speakers perceive the different meanings of *nägema* ‘to see’. The study provides evidence for the research questions Q1 and Q2 by investigating which meanings the verb *nägema* ‘to see’ has; and how language user’s perceive them. Q3 is answered by placing the results into a cross-linguistic context, and research question Q5 by applying both an experimental and a corpus-based method.

The second study takes the tactile experience perception verb *tundma* ‘to feel’ under observation with the help of two different experimental designs (a sorting task and a conceptual feature rating task). The goal of the study is to find out how Estonian speakers perceive the polysemy of the tactile perception verb *tundma* ‘to feel’, thereby providing evidence for the research questions Q1 and Q2. Study II provides evidence to research question Q3 by comparing the polysemy of the Estonian verb *tundma* ‘to feel’ to evidence from other languages.

Study III is an extensive corpus study dealing with the five experience perception verbs in Estonian. The goal of the study is to determine what types of usage patterns of perception verbs emerge from a corpus sample, and how this is tied to the polysemy of said verbs. Study III provides evidence to research questions Q1–Q4. Research question Q5 is relevant in all three studies. Throughout the studies, different methods are applied to polysemy research. The presentation of results is followed by an overall discussion in the next section.

3.1 Seeing the invisible (Study I)

The importance of vision, especially in Western culture, is hard to underestimate. Vision has been demonstrated to be the primary sensory modality, especially in Indo-European languages, but also in many other languages (e.g. Viberg, 1984; San Roque et al., 2014). Viberg (1984) argues for a hierarchy of the senses, with vision being the modality with the potential for most intrafield extensions (meaning a vision verb can potentially express all other modalities). The primacy of vision as a sense modality has also been shown for Australian languages by Evans and Wilkins (2000, p. 560), and a paper by San Roque et al. (2014) shows the dominance of vision in spoken language for 13 different languages around the world. Vision (together with audition) has also been shown to carry the most polysemous meanings (Storch and Aikhenvald, 2013a; San Roque et al., 2018) and has been shown to hold its prominence across different time periods in a diachronic study of English (Winter et al., 2018).

The Estonian *nägema* ‘to see’ is both frequent and polysemous in the corpus as shown in Section 1.3. As such, *nägema* ‘to see’ mirrors a picture that is seen

in many other languages. Study I offers an in-depth look at how the polysemy of *nägema* is perceived by native Estonian speakers as well as how it is reflected in the corpus. First, a sorting task was conducted. Sorting data of 20 participants was analysed with hierarchical cluster analysis (see Section 2.3.1). The dendrogram depicting the results from the analysis is presented in Figure 3.

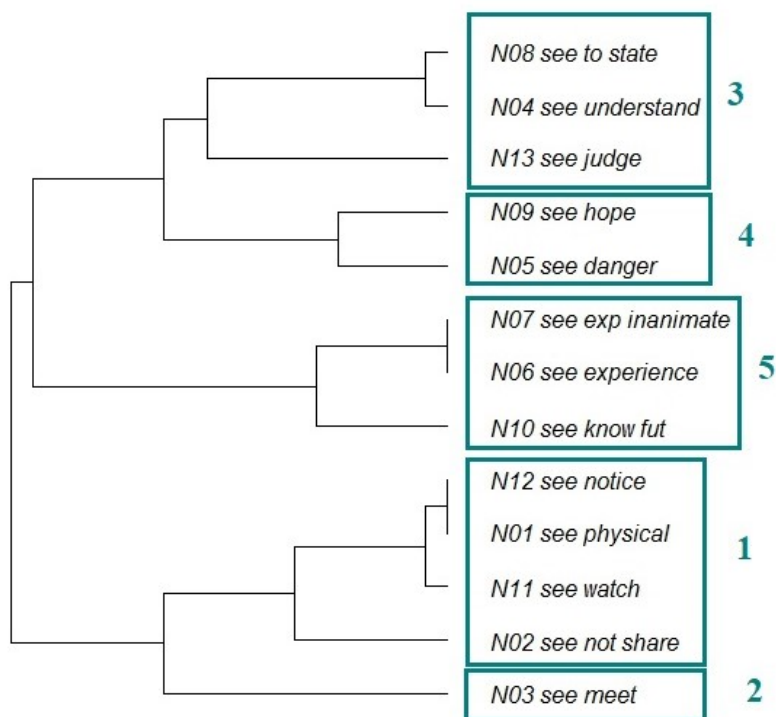


Figure 3: Meanings of *nägema* ‘to see’ on the basis of a sorting task, analysed with HCA (Proos, 2019b)

The meanings on *nägema* ‘to see’ group into five:

1. physically experiencing (e.g. N01 ‘see physical’, *näen seda maja* ‘I see this house’);
2. meeting someone (N03 ‘see meet’, *näeme varsti!* ‘see you soon!’);
3. understanding (e.g. N04 ‘see understand’, *ma näen, et muretsemine on mõttetu* ‘I see that it is pointless to worry’);
4. having a presentiment (e.g. N09 ‘see hope’, *kuidas näed Venemaa lähitulevikku?* ‘how do you see the near-future of Russia?’);

5. experiencing/going through something (e.g. N06 ‘see experience’, *olen sellega palju vaeva näinud* lit. ‘I have seen a lot of pain with this’ ‘I have worked very hard on this’).

To describe the emerged groups, the notion of image schemas (Lakoff and Johnson, 1980; Johnson, 1987) was used. The goal of this was to show that native speakers rely on underlying, conceptual similarities between the meanings when asked to complete a sorting task. For example, the group ‘*nägema* as understanding’ is described through the PATH image schema (Johnson, 1987, p. 114) and the group ‘*nägema* as experiencing’ through the FORCE image schema (Johnson, 1987, p. 43). This suggests that there are structural similarities between the different meanings of *nägema*, which enable speakers to organise the meanings in their mind.

The meaning extensions that *nägema* exhibits are in part similar to the ones we see in other European languages. For example, *nägema* represents the very common vision → cognition metaphor. However, this extension is not as productive as, for example, in English. Based on many European languages, Sweetser (1990) postulated that the vision → cognition extension is possible due to the similarities that knowing something has with vision, e.g. the relative objectivity of both experiences. In light of this, *nägema* ‘to see’ in Estonian has an uncommon meaning of ‘experiencing / going through something’. Mostly co-occurring with the noun *vaev* ‘pain, trouble’, *nägema* gets the meaning of going through an emotional (and to an extent physical) experience.

Next, a behavioural profile was constructed to look at the polysemy of *nägema*. Results from the corpus-based study show a different structure. The behavioural profile of *nägema* was based on a 700-sentence sample from the etTenTen corpus. Altogether 89 morpho-syntactic and semantic characteristics were annotated. Similarly to the results of the sorting task, the profile was analysed with hierarchical clustering. This offers a great opportunity to compare the two sets of results. The resulting dendrogram can be seen in Figure 4.

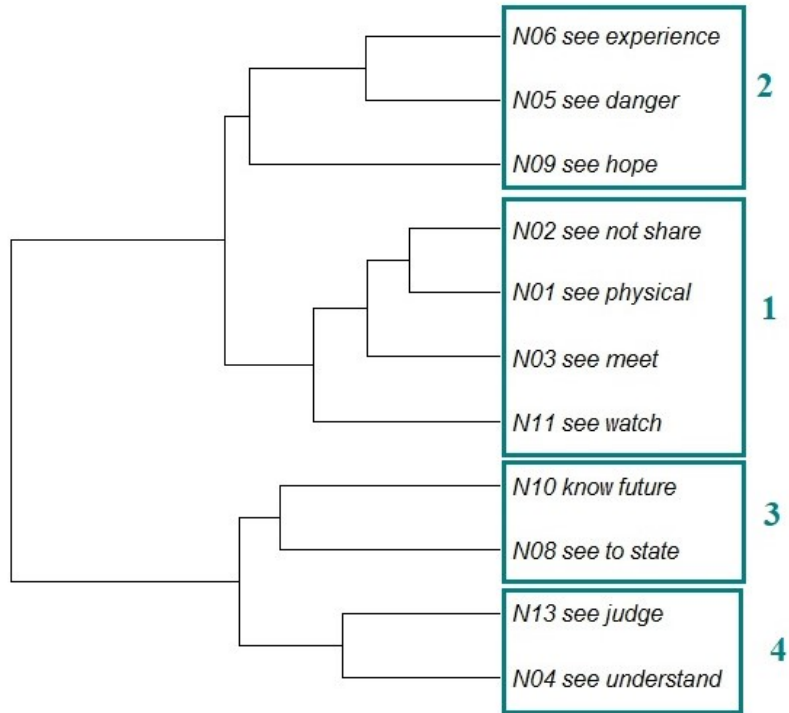


Figure 4: Meanings of *nägema* ‘to see’ on the basis of a behavioural profile analysis, analysed with HCA (Proos, 2019b)

The meanings of *nägema* ‘to see’ clustered into four groups; the driving factor behind this division was the nature of the goal of seeing. As such, the groups that emerged were the following:

1. *nägema* with a concrete bounded goal (e.g. N01 ‘see physical’, *näen maja* ‘I see a house’);
2. *nägema* with an abstract bounded goal (e.g. N06 ‘see experience’, *näen vaeva* ‘I see pain’);
3. construction-specific meanings (e.g. N08 ‘see to state’, *näe, siin ta ongi!* ‘look, there (s)he is!’);
4. *nägema* with an un-bounded goal (e.g. N04 ‘see understand’, *nägin, et sellest ei saa asja* ‘I saw that this would not work out’).

The main result of Study I was that when asked to sort sentences according to the meaning of the verb in the sentence, participants rely on underlying similarities between the different meanings. It was also shown that the vision → cognition extension that is so common across Indo-European languages is also present in

Estonian, although the extension is not as productive. In addition, a more uncommon meaning of *nägema* ‘to see’ – ‘experiencing’ was shown to be present in Estonian.

Across the two methods, the meanings relating to physical perception grouped similarly, while other meanings differed more depending on the methods that was used. Participants in the sorting task relied on conceptual similarities between the different meanings of *nägema* ‘to see’. For example, in the sorting task, meanings N08 ‘see to state’ and N10 ‘see know future’ were sorted into separate clusters, since the underlying conceptual patterns of these two meanings are different. However, since these are both construction-specific meanings, they were sorted into the same group on the basis of the BP analysis.

Study I concluded that the results about the polysemy structure of *nägema* ‘to see’ are dependent on the method that was used. It is difficult to replicate language users’ sorting decisions, since some of the perceived characteristics that language users base their similarity assessments on are impossible to ‘translate’ into objective, annotatable characteristics. This means that a corpus-based analysis might not resemble the analysis that is based on language users’ intuitions by way of an experiment. However, the structure that emerges from a corpus-based approach is also informative about the polysemy of *nägema* ‘to see’, especially in regards to its morphosyntactic usage patterns.

3.2 Feeling your neighbour and the clock (Study II)

As was shown in Section 1.3, tactile perception verbs exhibit a high degree of polysemy in Estonian. In the context of European languages, tactile perception in itself and its expression in language has not received a lot of attention (except Viberg, 2005, 2015) as it has been overshadowed by vision and audition verbs.

Study II takes a deeper look into the nature of the polysemy of *tundma* ‘to feel’. While cross-linguistically more common meanings like ‘feeling emotions’ do exist in Estonian, the Estonian *tundma* also exhibits more uncommon meanings like ‘knowing someone thoroughly’. First, a free pile sorting task was conducted with 66 participants. The goal of the sorting task was to find out, how Estonian native speakers perceive the meanings of *tundma* ‘to feel’. The results were analysed using multidimensional scaling. The resulting plot is presented in Figure 5. While in Study I the chosen method for analysing the sorting task results was HCA, in Study II I opted to use MDS as the analysis method. Both HCA and MDS visualise the similarities between objects in the data, however, the resulting visualisation from MDS does not show distinct groups of objects. Rather, it offers a concentration-cloud like structure, which can be argued to better represent the structure of polysemy. HCA, however, offers a clearer and in a way more strict picture of patterns in the data, which can be useful if we want to characterise the resulting groups of objects in a more concise manner. As one of the goals of

the dissertation is to map different methods of polysemy research, both HCA and MDS were used to showcase the potential differences in results.

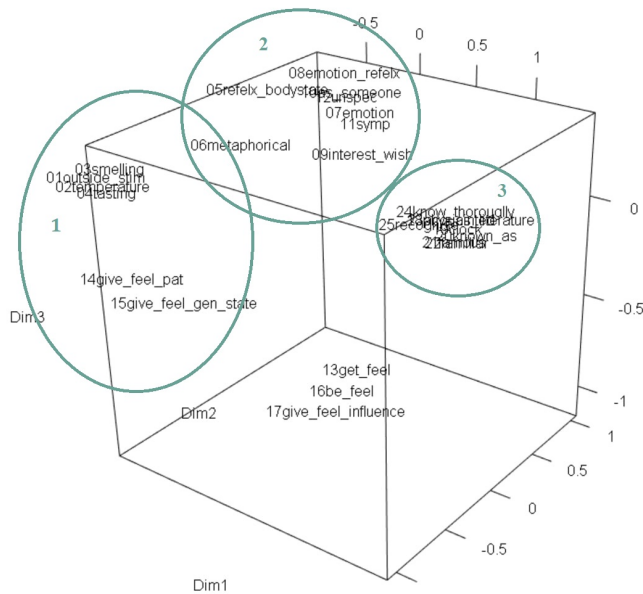


Figure 5: Multidimensional scaling plot of meanings of *tundma* ‘to feel’ based on a sorting task (Proos, 2020c)

The results showed that 25 separate meanings of *tundma* ‘to feel’, represented by sentences, formed three distinct meaning groups: 1) physical experiences (e.g. *tundsin kõhus valu* ‘I felt pain in my stomach’, *tundsin, et toas oli külm* ‘I felt that it was cold in the room’); 2) emotion-related meanings (e.g. *selle üle tunneb ta vaid rõõmu* ‘(s)he just feels happy about this’, *tunnen end turvaliselt* ‘I feel (myself) safe’); and 3) cognition-related meanings (e.g. *tunneb puid-põõsaid põhjalikult* lit. ‘feels trees-bushes thoroughly’ ‘knows a lot about trees and bushes’, *tunnen teda koolipõlvest saati* lit. ‘I feel him/her since school’ ‘I know him/her since school’).

There were also some meanings that did not fit any of the groups very well. These were very general meanings and constructions where *tundma* ‘to feel’ formed a compound verb with another verb (e.g. *tunda saama* ‘feel.inf1 get.inf2’). As is visible from Figure 5, these meanings group together in the lower part of the plot, around the zero point of the plot. Items that have a near-zero value are not well suited to any groups, i.e. they were not consistently sorted by the participants. Thus, they group together not because they were sorted together by the participants, but rather because they were sorted differently by the participants. It is thus not reasonable to consider these as a separate group.

It was concluded from the results of the sorting task that language users per-

ceive the cognition-related meanings as separate from the other meaning groups. Figure 5 shows that the emotion-related meanings and the physical meanings are divided on the first dimension (Dim1) from the cognition-related meanings. It was shown that this can be understood as the physical-metaphorical distinction. This means that while emotion-related meanings and physical meanings were still perceived as tied to the physical tactile sense (i.e. on the physical side of the dimension), the cognition-related meanings were perceived as separate from physical perception.

As the meaning extension from tactile perception to cognition is not a common one, a second experiment was conducted in Study II to take a closer look at the cognition-related meanings *tundma* ‘to feel’ carries. The goal of the study was to see whether the cognition-related meanings of *tundma* carry stronger physical characteristics than other Estonian cognition verbs such as *teadma* ‘to know’ or *aru saama* ‘to understand’. Following the theory of cognitive semantics, meaning extensions are possible only because some characteristics are carried over to the new abstract concept, i.e. there needs to be some similarity. Since *tundma* ‘to feel’ is a tactile perception verb, it was hypothesised that the cognition related meanings also carry some aspects of this physical meaning.

A conceptual feature rating task (see Section 2.3) was conducted, where native speakers had to rate both sentences with *tundma* ‘to feel’ in its cognition meaning and sentences with other cognition verbs like *teadma* ‘to know’ or *aru saama* ‘to understand’ on a variety of cognitive dimensions. It was expected that the ratings would differ. However, no statistically significant differences in ratings surfaced – the two types of sentences were assigned similar ratings.

This led to the conclusion that the meaning network of *tundma* ‘to feel’ is not simple – it does not have one, prototypical, physical meaning in the centre with its polysemous meanings stemming from it. It seems more likely that *tundma* ‘to feel’ is not a primarily physical verb at all. Instead, it has a more general meaning of proximal perception. Firstly, *tundma* ‘to feel’ lexicalises the experiences of smelling and tasting in addition to feeling. Touch, olfaction and gustation are considered to be proximal senses in comparison to vision and audition, which are distal senses. Secondly, the emotion-related and cognition-related meanings can also be argued to express proximal perception. The type of knowledge that is expressed by cognition-related meanings is experiential knowledge, meaning that one has to have had personal experiences and/or contact with the object of knowledge. Thus, there is ample evidence suggesting that *tundma* ‘to feel’ in Estonian is a general proximal perception verb.

3.3 Patterns that unify and identify (Study III)

Study III set off to look at the experience perception verbs (*nägema* ‘to see’, *kuulma* ‘to hear’, *tundma* ‘to feel’, *maitset tundma* ‘to taste’, *lõhna tundma* ‘to

smell’) together. The aims of Study III were 1) to map usage patterns of the experience perception verbs in a corpus sample; 2) to determine which contextual elements would best enable the speaker to derive the meaning of the verb in the sentence. The driving theory behind this approach is the distributional hypothesis, according to which the context that lexical units occur in is informative of their function (Harris, 1954). This principle has been shown at work on the lexical (Peirsman et al., 2010; Franco et al., 2019), morphosyntactical (Bresnan et al., 2007; Bresnan and Ford, 2010; Klavan, 2017; Grafmiller and Szmrecsanyi, 2018; Szmrecsanyi et al., 2019), and semantic (Janda and Solovyev, 2009; Divjak, 2015) level. Thus, looking at the usage patterns of perception verbs should enable us to gain knowledge about their meaning.

500 sentences per verb were collected from the etTenTen corpus with the exception of *maitset tundma* ‘to taste’, where 418 sentences remained after manually checking the sample, resulting in a full set of 2,418 sentences. This sample was then annotated for 9 different variables, concerning both morphosyntax as well as semantics. The annotation of the semantic variables followed a bottom-up approach where there were no *a priori* variable levels.

To determine the most significant contextual predictors, a random forest model (see Section 2.3.3) was fitted. The results can be seen from Figure 6.

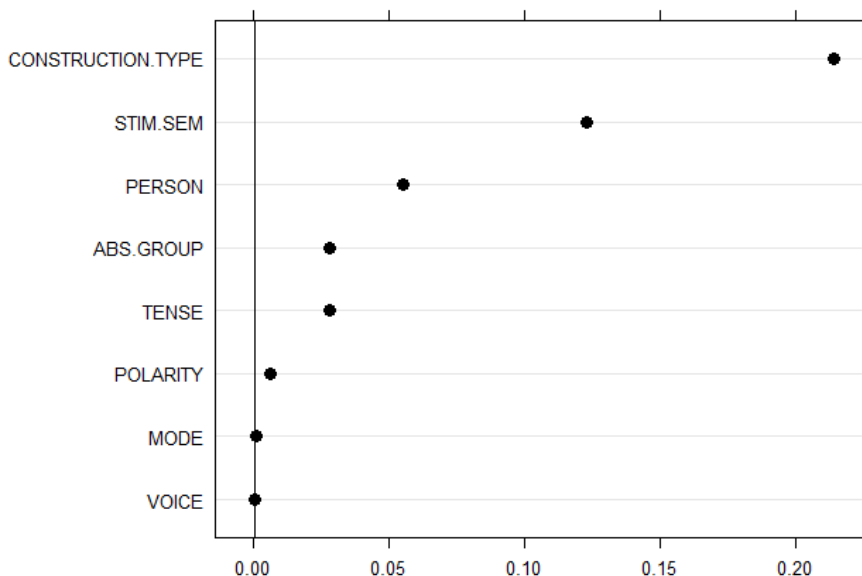


Figure 6: Variable importance plot (Proos, 2020d)

As Figure 6 shows, the construction type was the most significant predictor. This means that the type of construction that the perception verb occurs in is the

most indicative of the meaning of the verb itself. The semantic type of stimulus and the morphological person followed the construction type in importance. Analysing the combinations of these variables, Study III shows that some emerged usage patterns characterise the set of verbs as similarly behaving verbs, while other patterns are characteristic of a specific verb or even a specific meaning of that verb. A few examples are discussed.

The verbs exhibit different patterns of encoding the perception stimulus, i.e. what is being perceived. For example, in (1) the stimulus *teda* ‘him’ is encoded as the syntactic object, but in (2), the stimulus is encoded as a nominal modifier, *kanarbiku* ‘heather’.

- (1) Sookilpkonn peitu-vat mutta iga pisema krabina peale,
 bog.turtle.nom hide-quot mud.ill each tiny.comp.gen rustle.gen onto
 sellepärast on-gi **teda** peaaegu võimatu näha.
 because.of.this be-clitic **he.part** almost impossible see.infl
 ‘The bog turtle is said to hide in the mud for even a small rustle, that is why it is almost impossible to see **him**’
- (2) **kanarbiku** lõhn vahel juba paari meetri kauguselt
heather.gen smell sometimes already a.few.gen meter.gen away
 tunda!
 feel.infl
 ‘One can sometimes smell the **heather** even from a few meters away!’

The simple verbs *nägema* ‘to see’, *kuulma* ‘to hear’ and *tundma* ‘to feel’ exhibit a usage pattern, where the stimulus is classified as the semantic type ‘concrete’ and the verb occurs in an extended meaning. These three verbs can occur in a polysemous meaning that encodes a type of understanding or analysis of things. Each verb encodes a slightly different type of cognition. For example, *nägema* ‘to see’ has a polysemous meaning that encodes objective, analytical knowledge. *Tundma* ‘to feel’, however, has a meaning where the verb encodes personal, experiential knowledge.

When occurring with a stimulus that is classified as ‘abstract’, all of the verbs except *kuulma* ‘to hear’ have extended meanings that expresses an emotional experience. Again, the verbs determine the conceptual content of this experience. For example, *maitset tundma* ‘to taste’ is not very productive, but it is used to express the anticipation of something, e.g. *tunnen juba võidu maitset* ‘I can already taste the victory’. Here, ‘victory’ is paralleled with ‘food’, and the physical anticipation of tasting something good is paralleled to anticipating winning. *Tundma* ‘to feel’ is most productive with these types of stimuli, since ‘experiencing emo-

tions' is an attested polysemous meaning of *tundma* 'to feel'.

Some of the usage patterns are characteristic of only one verb or one meaning of a verb. For example, in Estonian, *kuulma* 'to hear' does not exhibit a very extensive meaning network. However, the few different meanings it does carry are rather entrenched in Estonian and even function as fixed expressions. When *kuulma* 'to hear' occurs with a noun phrase in the elative case, the meaning shifts to 'reported knowledge' or 'hearsay'. *Tundma* 'to feel' is also characterised by some specific constructions. For example, *tundma* 'to feel' is the only verb that occurs in a reflexive construction.

Study III showed how varied the usage patterns of Estonian experience perception verbs are. The study concluded that there are some patterns that characterise the class of verbs as a unit, but others that are predictive of a certain verb or a set of verbs. The key components to the variation of the usage patterns are both morphosyntactic and semantic. Perception verb polysemy has a significant effect on the usage patterns and *vice versa*. Although no direct causal link was established between meaning and form, they are inherently connected, especially in a language with rich morphosyntax like Estonian.

3.4 Discussion

The research questions this dissertation answers are as follows:

1. what are the various meanings that Estonian experience perception verbs carry? (Q1);
2. how is the polysemy of experience perception verbs conceptualised in the language user's mind? (Q2);
3. how does Estonian experience perception verb polysemy relate to cross-linguistic findings about experience perception verbs and their polysemy? (Q3);
4. which usage patterns are characteristic to Estonian experience perception verbs? (Q4);
5. which considerations need to be addressed in regards to the methods of polysemy research? (Q5).

In the first section of the Discussion, results in regard to research questions Q1–Q4 are discussed. Research question Q5 is discussed in the second section of the Discussion.

3.4.1 Estonian experience perception verb polysemy and cross-linguistic considerations

To answer research questions Q1–Q4, in what follows, the results presented in Section 3 are discussed. The results concerning all five sense modalities are discussed, and special consideration is given to how Estonian perception verbs compare to what has been found by previous research about other languages as well as to patterns that make the Estonian system stand out.

It was shown in Section 1.3 that *nägema* ‘to see’ is the most frequent perception verb in Estonian. To understand how Estonian native speakers perceive the verb’s polysemy, a sorting task was conducted (Study I). In many aspects, *nägema* ‘to see’ exhibits similar meaning patterns to those found in other Indo-European languages such as English (Sweetser, 1990; Winter et al., 2018), Swedish (Sjöström, 1999), and Lithuanian (Usoniene, 2001). For example, it is also used to express ‘understanding something’, although the meaning is not as productive as in, for example, English. Sweetser (1990) has noted that vision is often understood as an objective sense and thus it makes sense that the verbs of visual perception are used to conceptualise other experiences that are rational or objective in nature, i.e. tied to our intelligence as opposed to emotions, such as knowing or understanding.

However, Estonian was shown to also have a group of meanings not so common across the world’s languages. In Estonian, *nägema* ‘to see’ can also mean ‘experience / go through something’, as in example (3).

- (3) Vaese-d inimese-d saa-vad ikka **päevi** **näha**.
poor-pl person-pl get-3pl particle day.pl.part see.inf1
‘Poor people often go through difficult times (lit. get to see some days)’

In contrast to the pattern observed by Sweetser (1990) and others, the polysemy exhibited in example (3) is a generalisation of the visual perception meaning. The lexicalisation of knowing or understanding through vision verbs is driven by the objectivity of both experiences. On the contrary, in the case of (3), instead of specifying one aspect of vision and using that as the basis for a meaning extension, it seems the broader experience is taken as the basis.

It is worth noting that the ‘experience’ meaning of *nägema* ‘to see’ is not fully productive. There seems to be a fixed set of noun phrases that occur in this meaning, including nouns like *päevi* ‘days.pl.part’ or *vaeva* ‘pain.part’. In example (3), the noun phrase *päevi* ‘days’ is not negative in itself, but in this type of context, it is interpreted as ‘days of hardship’. Thus, it is mostly negative experiences that are expressed with this meaning.

Interestingly, the generalisation pattern is also what drives another type of ‘experiencing’ that is lexicalised by *nägema* ‘to see’. When the agent of the action

of seeing is inanimate, the meaning of the verb is also generalised. This is exemplified in (4). In the example, the verb is used to describe the state of the house, namely that it is an old building that has been in better and in worse state. The language users did not differentiate between (3) and (4), sorting them into the same pile. This suggests that they are perceived similarly, as a general ‘going through a negative experience’ meaning. Possibly, *maja* ‘house’ is simply personified in this context¹⁰, in which case (4) is a figurative use of the ‘experience’ meaning (exemplified in (3)).

- (4) *maja oli näinud parema-i-d ja halvema-i-d aegu*
 house be.3sg.pst see.app better-pl-part and worse-pl-part time.pl.part
 ‘the house had seen better and worse times’

While *nägema* ‘to see’ is the most frequent experience perception verb in Estonian, the tactile perception verb *tundma* ‘to feel’ is the most polysemous (see Section 1.3), contrary to what has been found for many other languages (San Roque et al., 2018). Verbs of tactile perception have not been found to carry a lot of polysemous meanings in previous research, especially in comparison to verbs of seeing and hearing. However, some re-occurring patterns have been found. For example, a very common strategy of languages is to express feeling emotions with a tactile perception verb (Ibarretxe-Antuñano, 2006; Sweetser, 1990; Viberg, 1984). Extensions into partaking in food, affecting, and tackling an issue have also been found (Ibarretxe-Antuñano, 2006). In addition, Viberg (1984, p. 143) has observed some intrafield extensions – cases where the vocabulary for one sense modality is also used to lexicalise another sense modality. In Swedish, Finnish, Estonian, Hungarian, Russian, Polish, Serbo-Croat (*sic*), and Bulgarian, the tactile perception verb combines into a compound verb with the nouns ‘smell’ and ‘taste’ to express olfactory and gustatory experiences (e.g. *lõhna tundma*, smell-part feel.inf2, ‘to smell something’ in Estonian) (Viberg, 1984, p. 143).

The results from the sorting task and the conceptual feature rating task (Study II) suggest that *tundma* might not be a strictly physical verb in the sense that its primary meaning is not ‘tactile experience’. Rather, it seems to be better described as a verb of proximal perception, encompassing different experiences like smelling, tasting, feeling emotions, and experiential knowledge. All of these experiences have in common a close and personal contact with the object of perception. The fact that this verb has a general meaning might also shed some more light on the frequency information about perception vs non-perception meanings in a corpus sample, presented in Section 1.3. It was shown that tactile verbs have the least amount (11%) of unambiguous physical perception meanings in a corpus sample compared to verbs of other perception modalities. One reason for this

¹⁰ It is worth mentioning that English also has a similar type of expression, e.g. *The car had seen better days*. Thus, to an extent, English exhibits this type of meaning extension as well for *to see*.

can be that *tundma* ‘to feel’ can be described as only having a physical perception meaning in a limited number of contexts, and is likely to have a more general meaning of proximal perception in most contexts.

Interestingly, the equivalents of *tundma* ‘to feel’ – the Swedish *känna* (Viberg, 2015) and Finnish *tuntea* (Viberg, 2005) – follow a similar pattern. According to Viberg (2005), in these languages, the verbs have a rather similar profile: they are also used to express a thorough knowledge of something or someone, as well as various sorts of emotions, both primary and social ones (Viberg, 2015). It seems then this is something common to the languages of the area. The connection to the area is especially interesting since while Finnish and Estonian are Finno-Ugric languages, Swedish is a Germanic language. In fact, we see the same root for the Swedish *känna* in German *kennen* – however, in German, the verb has no perception meaning. To speculate over the reasons behind these similarities, be they cultural or language contact-related, is beyond the scope of this dissertation. However, it does seem that at least a part of the Circum-Baltic languages exhibit this connection between tactile perception and general, proximal perception.

The semantics of *tundma* ‘to feel’ might also be explained by the etymology of the verb. Etymologically, *tund(t)-* is an old Uralic root (before 4000 BCE) and has cognates even in the genetically farthest languages like Nenets, Enets or Nganasan (Metsmägi et al., 2012). According to Metsmägi et al. (2012), the cognates in other Uralic languages exhibit the same meaning (extension), where the corresponding verb means ‘to know something or someone thoroughly’, ‘to recognize’. Furthermore, in some languages, the physical meaning is not attested (Metsmägi et al., 2012). Viberg (2015, p. 105–106) has shown that in Swedish, the meaning extension of *känna* ‘to feel, to know’ has been from cognition to perception. Considering that *tundma* ‘to feel’ cognates in some Finno-Ugric languages do not carry the perception meaning, and the Swedish etymology pointed out by Viberg (2015), it is most likely that *tundma* has also historically had the cognition meaning as its primary meaning, and the perception meaning is actually the extended meaning.

In a cross-linguistic context, it is not uncommon for perception verbs to have a more general meaning that is specified in usage. For example, van Putten (2020) shows that in Avatime, a Kwa language (Niger-Congo) spoken in Ghana, there is a perception verb *nu* that encodes all non visual perception (van Putten, 2020, p. 426). In Romance languages, the verb *sentir(e)* exhibits similar trends, potentially encoding hearing, touching, smelling, tasting and a variety of non-perceptual experiences (Enghels and Jansegers, 2013).

Contrary to what is shown to be true for a wide array of languages across the world by San Roque et al. (2014, p. 50), the audition verbs in Estonian do not carry a lot of meanings nor are they very frequent. The small number of polysemous meanings for the experience audition verb is in concordance with the small degree of variation in its usage patterns, especially compared to *nägema* ‘to see’ or *tundma* ‘to feel’ (Study III).

Most of the polysemous meanings that *kuulma* ‘to hear’ exhibits, are tied to a specific construction. Namely, when *kuulma* ‘to hear’ takes an object in the elative case, the meaning shifts to ‘reported knowledge / hearsay’. This meaning is also reported by Vanhove (2008, p. 348–349) as ‘learn’ and ‘know the story of’ in English. *Kuulma* ‘to hear’ can also function as a discourse particle. San Roque et al. (2018, p. 388–389) observed that English, Siwu, Spanish and Italian all used hearing verbs in a discourse function in relation to upcoming speech. This is also found in Estonian, where *kuulma* ‘to hear’ in the second person singular or plural imperative (*kuule*, *kuulge*) is also used in a discourse function. In addition, Estonian also uses *nägema* ‘to see’ in a discourse function (*näed* (2sg), *näe* (2sg.imp), *näete* (2pl)), as means of drawing the conversation’s partner attention to a topic.

As in so many other European languages, the vocabulary for smelling and tasting does not function as a productive source for meaning extensions in Estonian. These verbs have very similar usage patterns (Study III), meaning there are very little characteristics in usage patterns that are indicative of only one of the two verbs. The fact that both *maitset tundma* ‘to taste’ and *lõhna tundma* ‘to smell’ are compound verbs has a significant impact on how these verbs present in a corpus sample. Because the typical syntactic object ‘slot’ is already filled by the nouns *maitse-t* ‘taste-part’ and *lõhna* ‘smell.part’, the types of constructions they can occur in are severely limited. The constructions they do occur in, are shared by both of the verbs, making their behaviour in the corpus very similar. For example, to express the source of a taste/smell the syntactic object is modified by a noun; to express the quality of the taste/smell, the syntactic object is modified by an adjective. The fact that the verbs *maitset tundma* and *lõhna tundma* pattern together in this way can also be seen to be indicative of the nature of the senses themselves. Olfaction and gustation are physically connected sensations and classified together as chemical senses. The receptors in our oral and nasal cavities deal with both olfactory and gustatory stimuli (Bossomaier, 2012, p. 238). For example, as most of us have experienced, having a bad cold seriously affects our sense of taste as well. Thus, *maitset tundma* ‘to taste’ and *lõhna tundma* ‘to smell’ lexicalize senses that are rather intertwined.

The semantic range of the stimuli one can taste or smell is also not very broad, suggesting that these verbs do not exhibit a lot of meaning extensions. It is more the case that the extended meanings these verbs exhibit are fixed expressions such as *tunnen võidu maitset* ‘I (can) taste the victory’ or *tundsini tehingu juures raha lõhna* ‘I could smell the money on the deal’. The atypical meanings that *maitset tundma* ‘to taste’ and *lõhna tundma* ‘to smell’ occur in mostly refer to a brief, subtle contact with something or an expectation of something.

Some consideration has to be given to the question of whether to consider *maitset tundma* ‘to taste’ and *lõhna tundma* ‘to smell’ simply semantic extensions of the tactile perception verb *tundma* ‘to feel’ or if they should be considered to be

separate verbs. I consider both of these approaches to be true. On the one hand, we can indeed analyse the verb *tundma* ‘to feel’ as the focal point and we can state that this verb also lexicalises the sensory experiences of smelling and tasting – this is what was done in Study II. When looking at the possible meaning network of the verb *tundma* ‘to feel’, it makes sense to analyse smelling and tasting as a part of the meaning network – the verb has to have some sort of characteristics that allow it to express these types of experiences. Thus, to be able to analyse these characteristics, all variation should be taken into account.

On the other hand, it also makes sense to consider *maitset tundma* ‘to taste’ and *lõhna tundma* ‘to smell’ as separate verbs that lexicalise the experiences of tasting and smelling. If we approach perception vocabulary in Estonian from the question ‘Does Estonian have lexical means that express the experiences of tasting and smelling (and only those)?’ then these verbs ‘fill the slots’. Thus, if the aim is to get an overview of the whole system of experience perception verbs, how they are used, and which similarities/differences they exhibit as a class of verbs, it is reasonable to consider them as separate lexical units as was done in Study III.

This dissertation showed the variation of meanings and usage patterns Estonian experience perception verbs exhibit. Some of the meaning patterns coincide with previous research, done mostly on Indo-European languages. However, Estonian also exhibits patterns that unify an area encompassing Estonian, Finnish and Swedish. These results confirm yet again that our language of perception is extremely versatile and continues to provide insights to how we conceptualise the world around us.

However, attention has to be drawn to the fact that the results provided in this dissertation do not represent the whole range of Estonian perception language. It is entirely possible that by including other perception verbs (see Section 1.3), we would see different patterns emerge. That being said, this holds true for every language and every research. Nonetheless, by surfacing patterns for a specific set of verbs, we can firstly argue that what we are seeing at least holds true for a part of language, and is possibly indicative of a wider pattern; secondly, further research can be built upon this knowledge making results from a small sample an invaluable resource.

As a final note, based on the results presented in Sections 3.1–3.3, and in Section 1.3, I wish to touch upon the perception verb hierarchy presented and discussed primarily in Viberg (1984, 2001). Viberg presents a tentative perception verb hierarchy in his 1984 work, where he discusses how polysemy patterns of perception verbs tend to follow this hierarchy. In the hierarchy, vision has the topmost position, followed by hearing, touch, and smell and taste (the latter on an equal level) in that order.

The hierarchy is read in such a way that a verb used for one sensory modality has the potential to extend to also denote modalities to its right. Extension pattern from right to left is not possible. Furthermore, the extension pattern need not be

contiguous, i.e. a verb can “skip” a modality (Viberg, 1984, p. 147). In later works he (Viberg, 2001, 2005) also discusses that lexicalization patterns tend to follow this hierarchy, with sensory modalities on the left of the hierarchy being more likely to be lexicalized with simple verbs. Moreover, verbs on the left of the hierarchy tend to be more frequent and show more variation on the morphosyntactic level (Viberg, 2001, p. 1297).

I propose the the hierarchy presented in Figure 7 to hold for the Estonian perception verbs.

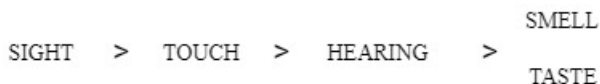


Figure 7: Estonian perception verb hierarchy

Vision is, without a doubt, at the top of the perception verb hierarchy in Estonian. Vision verbs are by far the most frequent, as shown in Section 1.3, the verb *nägema* ‘to see’ carries a lot of polysemous meanings (Study I) and shows large variation in a corpus sample (Study III). However, touch instead of hearing seems to be holding the second position in the hierarchy for Estonian. It was shown that tactile verbs are twice as frequent in the corpus in comparison to audition verbs. Moreover, Studies II and III show the large variation of both the semantic level as well as the morphosyntactic level for *tundma* ‘to feel’. Audition verbs fall short here, with little variation on the semantic and morphosyntactic level, especially in comparison to vision and tactile verbs.

As per the hierarchy suggested by Viberg (1984, 2001), touch here skips a modality and is extended to the modalities of smell and taste. Furthermore, the three leftmost verbs are all simple verbs, while the two rightmost verbs are compound verbs, again reflecting the original hierarchy. A caveat to the proposed hierarchy in Figure 7 is that it is mainly based on only experience perception verbs with some additional evidence from considering all the main perception verbs together. While this hierarchy will certainly need refinement, it is nonetheless an interesting viewpoint onto the system of Estonian perception verbs.

3.4.2 Methodological considerations

A discussion about methods of meaning research is especially necessary in the light of the cognitive semantic approach to meaning. Meaning in any case is something non-observable and subjective, and thus an extremely difficult subject to study. What makes it even more complicated is that meaning in the cognitive linguistic approach, in part, consists of encyclopaedic knowledge (e.g. Fillmore, 1985; Lakoff, 1987; Talmy, 2000), i.e. the knowledge and experiences we have of the world that surrounds us, and other concepts in this world. The question is

thus, how can we study all that goes into the nature of meaning; and what are we actually studying when using the methods we choose for meaning research?

Both corpus-based and experimental methods offer their own specific type of empirical evidence. Building one's research on corpus material is looking at how language is used, or rather, has been used in a specific setting. When looking at how meaning of a linguistic unit emerges from a corpus sample or is represented in it, we can only rely on the context that it is used in. However, this is one of the great advantages of a corpus-based method. If all language is language in use, then looking at a large number of uses can inform us about the language in question in a useful, theory-driven manner (Glynn, 2010; Glynn and Fischer, 2010). By quantifying the variation in usage, we can draw conclusions about re-occurring patterns. These in turn inform us about the function of the linguistic units occurring in these patterns (Gries and Divjak, 2010; Stefanowitsch, 2010). In other words, quantifying meaning allows us to get closer to what meaning is.

Linguistic experiments offer another type of evidence in meaning research. However, the conclusions drawn on the basis of experiments are not straightforward. Similarly to corpus-based methods, experimental methods do not offer a direct link into how meaning is organised in the language user's mind. In other words, we have access to language users' responses to our linguistic stimuli, but we cannot access what is the reason for these responses in a direct manner. The validity of the results and their generalisability has thus been discussed i.e. by Sandra and Rice (1995), Cuyckens et al. (1997), and Gibbs and Matlock (2001).

The main point of discussion is the question of whether what we are seeing from experimental results is actually representative of the speaker's mental lexicon. After all, the ultimate goal of empirical linguistics is to describe the cognitive reality, i.e. how language actually functions in a user's mind. For example, in the present dissertation, the sorting task method is used in Studies I and II. While the goal of the sorting task is to make conclusions about how different senses are organised in the language user's mind, it is questionable whether the results actually reflect the organisation of the senses in the mental lexicon (Sandra and Rice, 1995, p. 107–108).

While as outlined above, the types of evidence corpus-based and experimental methods offer are different, there are a number of other trade-offs between the methods (see e.g. Gilquin and Gries, 2009; Arppe et al., 2010). A clear advantage that corpus-based methods have over experimental methods is the amount of data that can be used to research a language phenomenon – corpus-based methods allow for a significantly larger amount of data. However, most corpus data is separated from the language use situation (with the exception of conversational or dialogue corpora). In written language corpora, it is extremely difficult and most often impossible to know the circumstances of the production of the data, which is often crucial to understanding the motivation behind usage.

Experiments, on the other hand, allow for control over the circumstances of

the data production and allow the researcher to manipulate them. This allows the researcher to gather data with less noise and ideally make sure that data necessary to answer the research question is indeed collected. What is gained in control, is lost in naturalness and the amount of data that can be gathered. Experimental settings are less natural environments for language use, and in these settings there is always a limit as to how much data participants can produce. Given the advantages and disadvantages of both an experimental approach and a corpus-based approach, it becomes crucial that a multitude of methods is applied to meaning research. Although it is evident that meaning as a cognitive phenomenon in a language user's mind cannot be directly accessed, by looking at how we react to stimuli and how we use language, we can place this new evidence in context with previously found evidence. This, in turn, enables us to piece together a possible representation of meaning as an integral part of language. In this process, corpus-based and experimental methods complement each other. In what follows, the methods used in this dissertation are analysed specifically in relation to polysemy research, with a focus on possible difficulties with interpreting the results.

As the corpus-based method, the behavioural profile method was used in Studies I and III. In Study I, concerning the perception verb *nägema* 'to see', the polysemous meanings were manually coded for every sentence along with a number of other semantic variables. As Section 3.1 showed, the results of the corpus-based portion of Study I were rather different from the experimental portion. In study I, a corpus-based approach still resulted in a by and large syntactic-behaviour-based grouping of the meanings, although there were more semantic variables and they were more detailed than the syntactic variables. In the sorting task in Study I, the sorting decisions were less based on syntactic behaviour, and more on underlying patterns of meaning characteristics. In addition, Study III, where all five experience perception verbs and their usage patterns were analysed, showed that construction variables were more significant in predicting the verb in the sentence than semantic variables.

Polysemy research with a behavioural profile method is not without its problems. The focus of the present research is the meaning of a verb, thus, the meaning functions as the dependent variable. We are looking at the different contexts that each meaning occurs in and trying to find patterns that would indicate one or the other meaning of the verb. However, this means that we need to assign meaning to the verb in every sentence in our sample. Glynn (2016) points out that this is problematic for a number of reasons. Firstly, meaning is our dependent variable, but by assigning meanings we are creating the dependent variable ourselves. Since meaning is dependent on context, we enter into a circle where we are influenced by the independent variables when we assign the meanings, and then look at whether these independent variables have an effect on our dependent variable.

Another complication that Glynn (2016) draws our attention to is the fact that this approach assumes that polysemous meanings are discrete units. However, if

we follow the principles of cognitive semantics as outlined in Section 1.1, this is not the case. It is more the case that meanings are categories with fuzzy boundaries. Thus, it is difficult to establish only one meaning per sentence. The results presented in Section 3.1 and Section 3.3 raise the question of the extent to which semantic information can be annotated in a corpus and how fruitful the endeavour is. The analysis of the usage patterns of the perception verbs (Study III) showed that verbs that exhibit less polysemy (*kuulma* ‘to hear’, *maitset tundma* ‘to taste’, *lõhna tundma* ‘to smell’) in comparison to highly polysemous verbs (*nägema* ‘to see’, *tundma* ‘to feel’) can be more sufficiently described in regards to their meaning variation.

The more polysemous the verb, the more morphosyntactic and semantic variation it exhibits, and *vice versa*. The variation for highly polysemous verbs can be very intricate and hard to pinpoint. For example, it is very difficult to annotate the difference between *nägema* ‘to see’ as watching (*olen seda filmi näinud* ‘I have seen this movie’) and *nägema* ‘to see’ as experiencing through vision (*kas näed seda hoonet?* ‘do you see that building?’). However, these two were occasionally sorted into different groups by language users, suggesting speakers do perceive them as different meanings.

Conversely, the difference between *maitset tundma* ‘to taste’ as physically experiencing taste and *maitset tundma* ‘to taste’ as anticipating something (*tundsin võidu maitset* ‘I could taste the victory’) is easier to describe. Annotating the stimulus is sufficient to showcase the different meanings: a stimulus that is perceivable via gustation is characteristic to the meaning ‘physically experiencing via taste’ and an abstract stimulus is characteristic to the ‘anticipating something’ meaning.

The polysemous meanings of verbs with a smaller degree of polysemy are more discrete, e.g. the polysemous meanings of *maitset tundma* ‘to taste’ are more discrete than the polysemous meanings of *nägema* ‘to see’. Language elements that are less polysemous are easier to differentiate and annotate than highly polysemous elements, where the meanings are much more intertwined. The research included in this dissertation illustrates how these differences have their effect on the method chosen for research. Corpus-based polysemy research is definitely fruitful and results in informative conclusions about the nature of the polysemy of the linguistic element in question, however, the extent to which inferences about polysemy can be made based solely on corpus information is highly dependent on the complexity of the polysemy.

Asking language users directly about how they perceive or use verbs, are a more direct window into the mind of the language user and offer an alternative to corpus-based methods. However, similarly to corpus-based methods, caution about the power of inferences related to actual information about meaning in the mind of a language user is advised. Offline designs in particular, like the ones used in the present dissertation, have to be approached with some reservations.

For example, a sorting task is used in Studies I and II. When sorting elements

into groups, it has to be kept in mind that the choice of elements in the task creates a kind of background system, in which the participants make their decisions. Thus, an item is sorted into a group if a participant "... finds any plausible rationale for relating it to prototypical members" (Langacker, 1987, p. 17). This means that the generality or granularity of the sorting items plays a huge role in the decisions of the participants. For example, when one is presented with a list of beings such as *cow*, *chicken*, *alien*, *human*, and *whale* one may be inclined to sort *cow*, *chicken*, *whale*, and *human* into one group. However, if the list of items only consisted of *cow*, *chicken*, *whale*, and *human*, one might sort *cow*, *chicken*, and *whale* into one group and *human* into a separate group. That is to say, sorting decisions are always contingent on the items being sorted, and inferences about the organisation of these items in the language user's mind cannot be considered permanent. In a similar manner, participants in Study II sorted 'smelling', 'tasting', 'feeling pain' and 'feeling temperature' into one group. This does not necessarily mean that these meanings are very similar at a conceptual level. They are, however, more similar to each other than to 'feeling emotions' or 'feeling physics'. The question of granularity is only one of the possible factors that can influence a participant's decisions.

In closing, researching meaning of any kind will always be difficult. We simply lack the methods today to say with full certainty what meaning even is, let alone draw conclusions about the intricacies of meaning networks in a language user's mind. We can, however, look at how language has been used by a large group of people in different settings by looking at corpus material. Finding patterns in a corpus can hint at decisions about language usage that many people share and allow us to estimate what is behind these decisions. On the other hand, by analysing how language speakers use or perceive language under specific circumstances, we can make contingent conclusions about a very small part of language and how it is organised in a person's mind. These results in turn can inform new theories, which can be tested in another set of specific circumstances, filling in yet another piece of the puzzle. Thus, no method should be left behind in pursuit of getting an optimally complete picture of the nature of meaning in language.

4 CONCLUSION

This dissertation offered an in-depth analysis of Estonian experience perception verbs – *nägema* ‘to see’, *kuulma* ‘to hear’, *tundma* ‘to feel’, *maitset tundma* ‘to taste’ and *lõhna tundma* ‘to smell’, and their polysemy. The nature of the polysemy of these verbs was researched with both experimental and corpus-based methods. The results of these studies confirm a cross-linguistically common trend of vision verbs being the most frequent in a language. It was shown that vision verbs are the most frequent in the corpus, with twice as many vision verbs observed than tactile verbs, which hold the second position frequency wise. The cross-linguistically observed meaning extension from vision to understanding is also present in Estonian, although it is not as productive as in many other European languages (Study I). The experience vision verb *nägema* ‘to see’ was shown to also carry a less frequent meaning in Estonian: going through an (emotional) experience.

Tactile perception verbs were shown to exhibit the most polysemous meanings in Estonian – a result that is not commonly observed cross-linguistically. The verb *tundma* ‘to feel’ was shown to have an extensive meaning network and Study II argues that this verb should be considered as a general proximal perception verb in Estonian, rather than an exclusively physical perception verb. The verb lexicalises different types of proximal experiences, ranging from the tactile meaning of e.g. feeling a touch on the skin and feeling internal experiences like emotions; to having experiential and/or personal knowledge of someone or something. The verb *tundma* ‘to feel’ also lexicalises the experiences of smelling and tasting. What ties all of these different experiences together is proximal contact with the stimulus, be it physical, mental, or emotional.

The Estonian *tundma* ‘to feel’ attests to the huge variety of ways humans can use language to conceptualise various experiences of the world. It echoes the nature of tactile perception itself. There are so many aspects to our sense of touch that it is not easy to pinpoint what this experience entails. A variety of experiences come together when we experience something by way of our sense of touch, both internal and external. Likewise, this seems to be the case in language as well. The richness of meaning of the Estonian *tundma* ‘to feel’ correlates with the richness of the sensory information that our sense of touch can give us.

Contrary to what is observed in previous research about other languages, the audition verbs in Estonian are not very frequent compared to vision verbs and tactile verbs. The meaning variation that the experience audition verb *kuulma* ‘to hear’ exhibits is rather small. The variation it does exhibit is largely tied to the construction it occurs in, as showed in Study III. When the stimulus of the perception is coded in the relative case (as opposed to the syntactic object cases partitive and genitive in Estonian), the verb acquires the meaning of ‘reported speech / hearsay’: a similar meaning has been observed for some Indo-European languages. In addition, *kuulma* ‘to hear’ functions as a discourse particle in the

2nd person singular and plural imperative form (*kuule, kuulge*). *Nägema* ‘to see’ also functions as a discourse particle in the second person singular in indicative and imperative, and in second person plural indicative (*näed, näe, näete*). This result confirms the cross-linguistically-attested pattern of perception verbs functioning as discourse particles. The development and nature of these particles has been investigated e.g. by Keevallik (2005, 2008).

The verbs for olfactory and gustatory perception are the least frequent in Estonian as well as the least varied in meaning. This result is in line with results from other European languages and confirms that Western culture tends not to put that much emphasis on the senses of smelling and tasting. The Estonian experience verbs of gustation and olfaction are compound verbs made up of the tactile experience verb *tundma* ‘to feel’ and a noun phrase lexicalising either taste or smell: *maitset tunda* ‘to taste’ and *lõhna tundma* ‘to smell’. Thus, the verbs are morphosyntactically different than the experience verbs for other modalities. The verbs exhibit similar semantic and morphosyntactic behaviour partly due to the similar morphosyntax. However, it is also likely that the lack of semantic variation is based on the low cultural importance that these verbs have in Estonian.

Based on an extensive corpus study (Study III), it was shown that polysemy of the verb plays a role in the usage patterns in which the verb can occur. Verbs that have a richer polysemy network such as *nägema* ‘to see’ and *tundma* ‘to feel’ also have more varied usage patterns. Form and meaning are inherently connected, and this dissertation demonstrated how establishing usage patterns can also shed light on the different meanings that a verb has.

This dissertation employed both experimental and corpus-based methods of polysemy research. It was demonstrated how these methods complement each other and how research into meaning demands triangulation of methods to get the fullest possible picture of this phenomenon. The dissertation provides examples of how the fruitfulness of corpus-based polysemy research is highly dependent on the complexity of the polysemy: the more complex the polysemy, the more caution should be taken in relying on corpus material. However, it was also shown that for verbs with a less complex polysemy, corpus-based methods offer insights not only about the polysemy of the verbs, but also about how meaning is inherently connected to form.

The experimental methods employed in this dissertation showed how language users perceive the polysemy of Estonian experience perception verbs and offered valuable insights as to how meaning might be represented in the language user’s mind. However, it was also discussed how experiment results are always contingent on the stimuli, and care has to be taken in interpreting the results as describing a permanent meaning structure. Overall, the dissertation showed how applying multiple methods can offer a more complete picture of polysemy than would be possible with any single method.

Although this dissertation offers a thorough look into the polysemy of exper-

ience perception verbs in Estonian, there is much research potential in investigating other Estonian perception verbs. Of particular interest would be the source-based perception verbs discussed in section 1.3. As shown, there is a large amount of variation here across all sense modalities. Especially interesting is the tactile source-based verb *tunduma* ‘to feel’. This a verb that can be used to encode experiences from most other sense modalities, making it a general verb. Future research could investigate, for example, which is the primary sense modality that this verb is used for by way of corpora; or what are the conditions under which this verb is used for expressing other sensory modalities.

The verb *tundma* ‘to feel’ also offers fruitful ground for further research. As was shown in this dissertation, the verb has a number of cognition-related meanings (e.g. *tunnen oma naabrit juba aastaid* ‘I have known (lit. ‘felt’) my neighbour for many years’). Study II showed that these cognition-related meanings were not rated differently than sentences with other cognition verbs like *teadma* ‘to know’ or *aru saama* ‘to understand’. However, *tundma* ‘to feel’ is not synonymous with the latter verbs. Thus, an investigation into the cognition-related meanings of *tundma* ‘to feel’ would be interesting. Specifically interesting would be looking into the type of knowledge that *tundma* ‘to feel’ expresses, considering the distinction between implicit, explicit, and tacit knowledge (Dienes and Perner, 1999; Ellis, 2005; Jung and Newen, 2010; Hays and Torres-Sahli, 2017).

Moreover, typological and anthropological work into the area-specific phenomenon of tactile perception verbs having cognition-related meanings would be very valuable. There is evidence to suggest that the tactile perception verb that is currently used to express cognition-related meanings did not extend semantically from the concrete field to the abstract; rather, the original meaning of this verb has been cognition-related, with the tactile perception meaning surfacing later. Research into the reasons behind this rather unusual meaning extension would widen our understanding of both the languages and cultures of a certain part of the Circum-Baltic area.

A topic that is not discussed in the present dissertation but, without a doubt, would be an interesting research topic on its own, is the use of Estonian perception verbs in spoken discourse. Throughout the dissertation, there is mention of perception verbs *nägema* ‘to see’ and *kuulma* ‘to hear’ also being used as discourse particles in certain constructions. Future research as to the variation and content of these particles in relation to similar particles in a cross-linguistic context as well as in relation to their perceptual content would replenish the knowledge we have about spoken language. More and more research is also focussing on spoken language in an effort to bring balance to the mostly written-language-based research tradition. Using Estonian corpora such as the Corpus of Spoken Estonian (Rääbis, 2013) and the Phonetic Corpus of Estonian (Lippus, 2013), investigations into the perception-based discourse particles would be a valuable contribution to both the research of Estonian and spoken language in general.

SUMMARY IN ESTONIAN

Eesti keele kogemustajuverbide tähendused ja kasutus

Keeltes kasutatakse tajumise väljendamiseks mitmeid erinevaid strateegiaid. Näiteks on keeli, milles leksikaalselt eristatakse ainult nägemist ja mitte-nägemist, st keeles puuduvad eraldi leksikaalsed vahendid kõikide tajude väljendamiseks. Sellised keeled on näiteks avatime (kvaat keel Nigeri-Kordofani hõimkonnas), duna (Üld-Uus-Guinea keel Paapua keelkonnas), itaalia ja hispaania (romaani keeled indoeuroopa keelkonnas), tselali (läänemaaja keel maaja keelkonnas) ja napuanmeni (Lõuna-Vanuatu keel Austroneesia hõimkonnas) keel (San Roque jt 2014: 41). Eesti ja inglise keeles eristatakse erinevate leksikaalsete üksustega aga kõiki viit taju. Samuti varieerub keeliti see, milliseid keelelisi vahendeid tajude väljendamiseks kasutatakse. Näiteks puuduvad enamikes Euroopa keeltes abstraktsed lõhnasõnad, selle asemel kasutatakse erinevaid konstruktsioone, näiteks “lõhnab nagu X”. On aga ka keeli, mille lõhnasõnavara on väga rikkalik, näiteks seri keel (Põhja-Ameerika isoleeritud indiaani keel) (O’Meara ja Majid 2020) või džahai keel (asli keel Austroaasia hõimkonnas) (Majid ja Burenhult 2014).

Väitekiri keskendub ühele vahendile, millega keeles tajukogemusi on võimalik väljendada – tajuverbidele. Tajuverbide all mõistetakse verbe, mis väljendavad mingi kogemuse saamist tajude kaudu, näiteks *nägema*, *kuulma*, *tundma*. Tajuverbid on tihti väga polüseemsed, st neil on mitu omavahel seotud tähendust. Enamasti on tähendus laienenud füüsilisest tähendusest abstraktseks – näiteks verb *nägema* tähendab nägemismeele abil kogemist, aga võimalik on ka *vaeva näha*, mis juhul *nägema* tähendus on laienenud emotsiooni kogemisele.

Kuigi spetsiifilisemad tähendusülekanded on tajuverbide puhul keeliti erinevad, leidub ka sellised üldiseid mustreid, mis ühendavad väga mitmeid keeli, nt on leitud, et tähendusülekanne taju → kognitsioon¹¹ on väga laialt levinud (Ibarretxe-Antuñano 2008). Samas erineb keeliti see, milline viiest tajust on selle ülekande produktiivne allikas. Inglise keeles ning ka paljudes teistes indoeuroopa keeltes väljendatakse kognitsiooniga seotud kogemusi eelkõige nägemist tähistavate sõnadega (Sweetser 1990: 33–34). Austraalia keeltes aga on allikaks kuulmistajuverbid (Evans ja Wilkins 2000). Rootsi, norra ja soome keeles on taktiilse (e kompimis-) taju verbid need, mis tihti kognitsiooniga seotud kogemusi väljendavad (Viberg 2005, 2015). On ka selliseid keeli, kus kognitsiooni väljendatakse lõhnataju sõnadega (Storch 2013).

¹¹ Inglise keeles räägitakse tähendusülekandest *perception* → *cognition*. Siinkohal aga on ingliskeelne *cognition* kitsam mõiste kui eestikeelne vaste ‘tunnetus’, mis sellisena hõlmab kõiki tunnetuslikke protsesse (sh aistinguid, taju, mälu jt). Kitsam *cognition* siin viitab ainult tunnetuse intellektuaalsele osale, st sellistele protsessidele nagu mõistmine. Siin ja edaspidi on eestikeelne ‘kognitsioon’ kasutatud just selles kitsamas määratluses.

Viberg (1984) on loonud tajuverbe keeltes organiseeriva süsteemi e põhitajuverbe paradigma. Selles süsteemis jagunevad verbid viie taju (nägemine, kuulmine, kompimine, maitsmine, haistmine) ning kolme dünaamilise süsteemi (tegevuse, kogemuse, allikapõhise) vahel, moodustades seega kokku viisteist potentsiaalset leksikaalset üksust. Tegevusverbi ja kogemusverbi erinevus on verbiga väljendatud tegevuse tahtlikkus – tegevusverbi puhul on agent aktiivne ning tajumine tahtlik ja suunatud (nt *vaatama*), kogemusverbi agent on pigem passiivne tajuja (nt *nägema*). Allikapõhise verbi agent on stiimuli rollis (nt *välja nägema*).

Väitekirjas keskendutakse kogemustajuverbidele *nägema*, *kuulma*, *tundma*, *maitses tundma* ja *lõhna tundma*, nende eri tähendustele ning kasutusmustritele eesti keeles ning võrdluses teiste maailma keeltega. Tähtsuse uurimisel on tähtis roll meetodi olemusel – kuidas uurida midagi, mida ei eksisteeri keelekasutaja meelest väljaspool? Seega keskendun töös ka tähenduse uurimise metodoloogilisele küljele.

Väitekirjas vastan järgmistele uurimisküsimustele.

- Millised tähendused on eesti keele kogemustajuverbidel? (K1)
- Kuidas tajuvad keelekasutajad kogemustajuverbide polüseemiat? (K2)
- Kuidas suhestub eesti keele kogemustajuverbide polüseemia maailma keelte kogemustajuverbide polüseemiaga? (K3)
- Millised kasutusmustrid on omased eesti keele kogemustajuverbidele? (K4)
- Millised metodoloogilised kaalutlused on tähenduse uurimisel olulised? (K5)

Väitekirja põhineb kolmel uurimisel (Proos 2019, edaspidi I uurimus, Proos 2020c, edaspidi II uurimus, Proos 2020d, edaspidi III uurimus). Esimese uurimisküsimusega (**K1**) tegeletakse läbivalt kõigis kolmes väitekirja kaasatud uurimuses. Küsimusele vastamiseks olen kasutanud nii katselisi meetodeid (I ja II uurimuses), korpuspõhiseid meetodeid (I ja III uurimuses) kui ka eesti keele sõnaraamatuid. Väitekirja eesmärk ei ole siiski pakkuda lõplikku nimekirja kõigi kogemustajuverbide kõigist tähendustest. Eelkõige on eesmärk kirjeldada, millist tüüpi tähendusi eesti keele kogemustajuverbidel on ning kuidas need väljenduvad. Töö esimene uurimisküsimus on oluline eelkõige eesti keele seisukohalt, pakkudes süvendatud infot eesti keele toimimise ja olemuse kohta.

Lisaks teadmisele, millised tähendused eesti keele kogemustajuverbidel on, pakub väitekirja sissevaadet ka sellesse, kuidas eesti keele kõnelejad kogemustajuverbe tajuvad (**K2**). Esimese ja teise uurimuse jaoks viisin läbi keelekatseid, mille tulemused võimaldavad heita pilku sellele, kuidas tähendused keelekasutaja meeles on organiseeritud. Kuigi väitekirjas kasutatud katsed ei peegelda päris otseselt seda, milline on tähenduste struktuur keelekasutaja meeles, siis meetodite kombineerimisel jõuame siiski sellele samm-sammult lähemale.

Kolmas uurimisküsimus (**K3**) asetab eesti keele kogemustajuverbid ja nende tähendused võrdlusesse teiste maailma keeltega. Sellele küsimusele vastavad läbivalt kõik väitekirja uurimused. Asetades eesti keele teiste keelte konteksti, on võimalik koguda teadmisi eesti keele eripärade kohta ning samal ajal laiemalt ka selle kohta, kuidas tajud üldiselt keeles saavad väljenduda. Seega panustab uurimisküsimus **K3** laiemalt ka inimese maailmataju uurimisse.

Neljas uurimisküsimus (**K4**) tegeleb kogemustajuverbide kasutusega ning seda käsitlen III uurimuses. Tajuverbide kasutuskontekstide uurimine võimaldab paremini aru saada, kuidas inimene tajuverbe mõistab.

Viimase uurimisküsimuse (**K5**) keskmes on keelelise tähenduse uurimise meetodika. Väitekirjas kasutan polüsemia uurimiseks mitut erinevat meetodit. Meetodite paljususe eesmärk on ühelt poolt saada võimalikult täielik pilt niivõrd raskesti tabatavast nähtusest kui seda on tähendus, aga teisalt ka panustada polüsemia uurimise metodoloogilisse arengusse. Sõnatähenduse selgitamiseks sobivate meetodite otsimise küsimusega tegelevad läbivalt kõik väitekirja uurimused. I ja III uurimuses kasutan korpuspõhiseid meetodeid ning I ja II katselisi meetodeid.

Teoreetilised lähtekohad

Töö teoreetiliseks lähtekohaks on kognitiivne semantika ning täpsemalt polüsemia käsitlus selles. Kognitiivse keeleteaduse põhiteese on, et keel on osa inimese kognitiivsest süsteemist, mitte eraldiseisev üksus. Seega on keel seotud inimese taju ja mootorikaga ning seekaudu ka kõigega, mida enda ümber ja sees kogeme (Langacker 1987a,b; Talmy 2000, 2003; Gibbs 2006). Polüsemia olemust mõistetakse kognitiivses semantikas kategoriaalsena ning prototüübi teooriast lähtuvalt (Rosch 1978, Geeraerts 2006). Polüsemise sõna tähendused moodustavad võrgustiku, mis jagab tähendused nii ühe kategooria sees kui ka kategooriate vahel perekondliku sarnasuse järgi. Igas kategoorias on prototüüpsemaid ja vähem prototüüpsemaid liikmeid ning kategooriate piirid on hägusad, mitte diskreetsed (Rosch ja Mervis 1975: 575).

Kognitiivne lähenemine järgib ka laiemalt kasutuspõhise lähenemise põhitõdesid (vaata nt Gries 2006; Gries ja Divjak 2009; Glynn ja Fischer 2010; Glynn ja Robinson 2014). Kasutuspõhised keeleteooriad lähtuvad eelkõige seisukohast, et kogu keel on keel kasutuses ning igal kasutusjuhul on oma funktsioon. Kontekst, kus konkreetne keelend esineb, annab edasi infot selle keelendi funktsiooni kohta. Seda põhimõtet tuntakse ka kui distributiivset hüpoteesi (Harris 1954). Kuna keelendi tähendus on üks selle funktsioonidest, siis on võimalik keelendi tähenduse kohta teha järeldusi selle põhjal, millises kontekstis keelend esineb (Gries ja Divjak 2010: 338). Väitekirjas järgitakse distributiivset hüpoteesi kui põhimõtet, mille järgi kasutuskontekst on tähenduse olemuslik osa.

Eesti keele põhitajuverbid

Järgides põhitajuverbide paradigmat (Viberg 1984) on tabelis 1 välja toodud eesti keele põhitajuverbide võimalik paradigma.

Tabel 1. Eesti keele põhitajuverbide võimalik paradigma

	Kogemus	Tegevus	Allikapõhine
Nägemistaju	nägema	vaatama	välja nägema
Kuulmistaju	kuulma	kuulama	kõlama
Kompimistaju	tundma	katsuma	tunduma
Maitsmistaju	maitset tundma		maitsuma
Lõhnataju	lõhna tundma	nuusutama	lõhnama

Et eesti keele tajuverbe paremini kirjeldada ning iseloomustamaks kogemustajuverbide kohta põhitajuverbide paradigmat, kasutasin sagedusinfot eesti keele veebitekstide korpusest etTenTen. Võrdlesin kõikide põhitajuverbide absoluutsagedusi korpuses. Sagedusinfo kogumise eesmärk oli ka võrrelda eesti keele tajuverbide sagedusi teiste maailma keelte tajuverbide sagedusmuustritega, st selgitada välja, kas eesti keele tajuverbide sagedusmuustrid peegeldavad varem teiste keelte kohta saadud tulemusi. Selgus, et absoluutsageduste poolest on esimesel kohal nägemistaju põhiverbid, mida oli korpuses kokku 649 968. Sellele järgnesid kompimistaju põhiverbid, mida leidis 361 257, kuulmistaju põhiverbide sagedus korpuses oli 172 335. Seega oli nägemistajuverbe pea kaks korda rohkem kui kompimistaju verbe, mida omakorda oli peaaegu 2 korda rohkem kui kuulmistaju verbe. Maitsmistaju ja lõhnataju põhiverbe oli korpuses teistega võrreldes väga vähe, vastavalt 11 831 ja 5663. Siiski on sellesse korpusvalimisse kaasatud vaid tabelis 1 ära toodud verbid, seega ei paku valimi põhjal tehtud statistika ülevaadet kõikide eesti keele tajuverbide sagedussuhetest.

Eesti keele seletava sõnaraamatu (EKSS, Langemets jt 2009) järgi on kõikidel põhitajuverbidel rohkem kui üks tähendus. Kõige rohkem, kokku 29 tähendust, on nägemistaju põhiverbidel. Samas kõige rohkem tähendusi ühe verbi kohta on kompimistaju kogemusverbil *tundma* – 18. Kompimistaju põhiverbidel kokku on EKSSi järgi 26 tähendust. Kuulmistaju põhiverbidel on juba vähem tähendusi, kokku 15. Verbil *maitsuma* on tähendusi 3, samuti lõhnataju verbil *nuusutama*. Lõhnataju allikapõhisel verbil *lõhnama* on 2 tähendust.

Illustreerimaks eri tähenduste jaotust korpuses, analüüsisin 3000-lauselise valimit, kus kõik põhitajuverbid olid esindatud 200 lausega (nt 200 lauset verbiga *nägema*, 200 verbiga *vaatama* ja 200 verbiga *välja nägema*). Valimis märgendasin iga lause puhul, kas tajuverb on lauses füüsilise taju tähendusega või mingi muu tähendusega tähendusega. Füüsilise tähendusena märgendasin vaid need laused, kus füüsilise tajumise tähendus oli eksplitsiitne. See tähendab, et kui verbi tähendust ei saanud üheselt määrata füüsiliseks, märgendasin selles lauses verbi

tähenduse mittefüüsiliseks. Kuigi valimis oli üldiselt füüsilisi tähendusi rohkem kui mittefüüsilisi, olid kompimistaju verbid siin silmatorkav erand. Kompimistaju põhiverbide esinemistest vaid 11% (n = 64) oli füüsilises tähenduses. Võrdluseks: maitsmis- ja haistmisverbide puhul moodustasid füüsilised tähendused vastavalt 87% (n = 523) ja 85% (n = 516). Seega näeme, et eesti keele põhitajuverbid erinevad üksteisest oluliselt, nii absoluutsageduste, tähenduste arvu kui ka tähenduste jagunemise poolest.

Väitekirja kolm uurimust keskenduvad eesti keele kogemustajuverbidele erinevas mahus ning erinevatest vaatepunktidest. I uurimuses käsitletakse verbi *nägema* polüseemiat katse- ja korpusandmete põhjal. II uurimuses kasutatakse kaht erinevat katset, et uurida tajuverbi *tundma* polüseemiat. III uurimuses käsitletakse kõigi kogemustaju põhiverbide kasutusmustrid korpusvalimis.

Materjal ja meetod

Väitekirja materjal koosneb kolme keelekatse tulemustest ning kahest korpusvalimist. Katseliste meetoditena on väitekirjas kasutatud sorteerimiskatset (I ja II uurimuses) ja kontseptuaalsete tunnuste hindamise katset (II uurimuses). Sorteerimiskatse kasutamise eesmärk oli teada saada, kuidas eesti keelt emakeelena kõnelejad tajuverbide polüseemiat tajuvad – milliseid sama verbi tähendusi tajutakse sarnasena, milliseid erinevana. Polüseemia uurimiseks on sorteerimiskatset kasutanud ka näiteks Sandra ja Rice (1995), Beitel jt (2001), Gibbs ja Matlock (2001) ja Raukko (2003). I uurimuses sorteerisid katseisikud lauseid, mis sisaldasid verbi *nägema* selle eri tähendustes. II uurimuses korrati katset, kasutades seekord verbi *tundma* eri tähendusi.

Kontseptuaalsete tunnuste hindamise katset (Troche jt 2014, 2017) kasutati ainult II uurimuses. Katse on algselt välja töötatud selleks, et uurida mõistete väärtusi abstraktsuse-konkreetsuse skaalal. Väitekirjas on kasutatud katse muudetud varianti, varem ei ole seda katsedisaini minule teadaolevalt polüseemia uurimiseks kasutatud. Katse eesmärk oli välja selgitada, kas lauseid, mis sisaldavad verbi *tundma* teadmiseiga seotud tähenduses, tajutakse teist moodi kui teisi teadmiseiga seotud verbe, nagu *teadma* või *aru saama*. Tabelis 2 on kõigi väitekirjas kasutatud katsete andmed.

Tabel 2. Väitekirjas kasutatud katsete andmed

	UURIMUS 1	UURIMUS 2	
	Sorteerimiskatse verbiga \textit{nägema}	Sorteerimiskatse verbiga \textit{tundma}	Kontseptuaalsete tunnuste hindamise katse
Katseisikute arv	20 (N 11, M 9)	66 (N 53, M 13)	68 (N 61, M 7)
Katseisikute vanus	keskmine 24 a vahemik 20–50 a	keskmine 31 a vahemik 21–65 a	keskmine 29 a vahemik 18–57 a
Katse stiimulite arv	lauseid: 13	lauseid: 25	lauseid: 72

Väitekirja korpusmaterjal koosneb kahest valimist. Valimite ülevaade on esitatud Tabelis 3. Korpusvalimit on kasutatud I ja III uurimuses. Mõlemas uurimuses rakendati käitumisprofili meetodit (Gries 2006, 2010), mis on polüsemia korpuspõhisel uurimisel häid tulemusi andnud (Berez ja Gries 2008, Gries ja Divjak 2009, Glynn 2014, Jansegers jt 2015, Jansegers ja Gries 2020). Käitumisprofili meetod põhineb kasutuspõhisel lähenemisel keelele. Selle lähenemise järgi saame teha järeldusi keelendi tähenduse kohta, vaadates seda, millises lausekontekstis see esineb. Valimis märgitakse iga lause puhul ära hulk tunnuseid, mis konkreetset kasutusjuhtu iseloomustavad, kusjuures tunnused võivad olla nii semantilised, morfosüntaktilised kui ka diskursuspõhised. Tunnuste esinemise sageduse põhjal luuakse tabel, mida analüüsitakse statistiliste meetoditega.

Tabel 3. Väitekirjas kasutatud korpusvalimid

	UURIMUS 1	UURIMUS 3
Korpus	Tasakaalus korpus (BCE)	Eesti veebikorpus (etTenTen)
Korpuse suurus	15 mln sõna	270 mln sõna
Valimi suurus	700 lauset	2418 lauset
Verbid	<i>nägema</i>	<i>nägema, kuulma, tundma maitset tundma, lõhna tundma</i>

Nii katse- kui ka korpusandmete analüüsimiseks kasutasin statistilisi meetodeid. I uurimuses kasutasin korpusandmete analüüsimiseks hierarhilist aglomeraatiivset klasteranalüüsi, mis on kirjeldav, andmeid korrastav meetod (Divjak ja Fieller 2014). III uurimuse korpusandmete analüüsiks kasutasin juhumetsa meetodit (Breiman 2001, Tagliamonte ja Baayen 2012), mis liigitub statistiliselt järeldavate meetodite alla. Juhumetsa meetodi eesmärk on välja selgitada uuritava keelendi varieerumist enim mõjutavad tunnused.

I uurimuses kasutatud sorteerimiskatse tulemuste analüüsiks kasutasin samuti hierarhilise klasteranalüüsi meetodit. II uurimuse sorteerimiskatse andmete analüüsil kasutasin multidimensionaalset skaleerimist (vt nt Levshina 2015), mille tööpõhimõte on sarnane hierarhilise klasteranalüüsiga, kuid visualiseerimise meetod on viimasest erinev. Kontseptuaalsete tunnuste hindamise katse analüüsiks II uurimuses kasutasin segamudeldamist, mis võimaldab ühes mudelis analüüsida nii fikseeritud kui ka juhuslikke faktoreid (vt nt Winter 2013, Gries 2013, Baayen 2014). Andmeid analüüsisin vabavaralise tarkvaraprogrammiga R (R Core Team 2019). Väitekirja andmestikud ja statistilise analüüsi skriptid on avaldatud ja kättesaadavad repositooriumis Open Science Framework (Proos 2019, 2020a, 2020b).

Tulemused ja arutelu

Väitekirjas esitatud põhitajuverbide sagedusandemte põhjal saab väita, et *nägema* on eesti keele kõige sagedasem põhitajuverb. Sellel verbil on väga rikkalik polüsemia – st palju erinevaid, üksteisega seotud tähendusi. I uurimuse eesmärk oli selgitada, kuidas eesti keelt emakeelena kõnelejad neid tähendusi tajuvad. Sor-teerimiskatse tulemused näitasid, et moodustunud tähendusrühmad peegeldavad tähendusmustreid, mis on sarnased ka paljudes teistes Euroopa keeltes, nt inglise keele (Sweetser 1990, Winter jt 2018), rootsi keele (Sjöström 1999) ja leedu keele põhitajuverbidel (Usoniene 2001). Näiteks kasutatakse ka eesti keele verbi *nägema*, et väljendada teatud tüüpi mõistmise ja arusaamise kogemusi (nt *Ma ei näe siin probleemi*), kuigi see tähendusülekanne pole nii produktiivne kui näiteks inglise keeles.

Kuid eesti keele verbil *nägema* on ka rühm tähendusi, mis maailma keelte kontekstis nii tavalised ei ole. Nimelt võib eesti keeles verbiga *nägema* väljendada ka isiklikku läbielamist/kogemist (nt *Vaesed inimesed saavad ikka päevi näha.*). Tihti on sellisel kasutusel juures negatiivne konnotatsioon, eelkõige väljendatakse nii negatiivseid kogemusi. Huvitav on ka see, et selles tähenduses võib agent, s.o nägija olla ka elutu, nt *maja oli näinud paremaid ja halvemaid aegu*. Teisalt võib neis näidetes agendi elusust tõlgendada ka personifikatsioonina.

Kõige polüseemsem kogemustajuverb ning põhitajuverb üldiselt on kompimistaju verb *tundma*. Võrreldes paljude teiste maailma keeltega (San Roque jt 2018), on selline tendents pigem haruldane – kompimistaju verbid on harva väga polüseemsed. Samas aga on leitud, et mõned tähendusmustrid korduvad paljudes maailma keeltes, nii näiteks väljendatakse kompimistaju verbidega väga tihti emotsioonide kogemist (Ibarretxe-Antuñano 2006, Sweetser 1990, Viberg 1984). On ka leitud, et kompimistaju verbidega saab mõnes keeles väljendada toidu maitsemist ning probleemi lahendamist (Ibarretxe-Antuñano 2006). Nii nagu eesti keelgi, kasutavad ka rootsi, soome, ungari, vene, bulgaaria, serbia ja horvaatia keel maitse ja lõhna tundmise kogemuse väljendamiseks kompimistaju verbi (Viberg 1984), millele lisatakse vastav nimisõna (Viberg 1984) (eesti keeles *maitset* või *lõhnatundma*).

Väitekirja II uurimus keskendus eesti keele kogemustajuverbile *tundma*. Selgus, et tegemist ei ole eelkõige füüsilise taju verbiga, pigem ongi *tundma* tähendus üldisem ning seda võib kirjeldada kui proksimaalse taju verbi. Kompimis-, haistmis- ja maitsemistaju liigitataksegi proksimaalsete tajude alla, vastandiks distaalsetele tajudele nagu nägemine ja kuulmine, kus tajuja ei pea olema tajutavaga otseses kontaktis. Verbiga *tundma* saab seega väljendada selliseid kogemusi nagu maitse tundmine, lõhna tundmine, temperatuuri tundmine, aga ka kogemuslikku teadmist (nt *tunnen oma naabrit hästi*). Neid kogemusi ühendab isiklik otsene kontakt sellega, mida kogetakse.

Niisiis on *tundma* üldise tähendusega verb, mille tähendus spetsifitseerub kasutuskontekstis. Maailma keelte lõikes ei ole selline tähendusmuster siiski ainu-

laadne. Näiteks avatime keeles väljendatakse tajuverbiga *nu* kõiki tajukogemusi peale nägemise(van Putten 2020: 426). Samuti on romaani keelte *sentir(e)* selline verb, millega saab väljendada nii kuulmist, puudutamist, haistmist kui ka maitsmist, aga ka mitmeid mitte-tajukogemusi (Enghels ja Jansegers 2013).

San Roque jt (2014: 50) 13 keele suulise andmestiku põhjal tehtud uurimuses olid kõige polüsemsemad nägemistaju verbid, neile järgnesid kuulmistaju verbid. Käesoleva väitekirja III uurimusest selgus, et eesti keele kuulmistaju verbidel eriti palju tähendusi ei ole. III uurimuse tulemused näitasid ka, et kogemustajuverbi *kuulma* kasutusmustrid varieeruvad võrreldes väga polüsemsete verbidega *nägema* ja *tundma* vähe. *Kuulma* tähendused on enamasti konstruktsioonispetsiifilised, st üks tähendus väljendub ainult ühes kindlas konstruktsioonis. Nii näiteks on juhul, kui *kuulma* objekti rollis esineb elatiivis nimisõnafraas, verbi tähendus lauses “vahendatud teadmine”(nt *Olen sellest mehest kuulnud*). “Vahendatud teadmise” tähendust esineb kuulmisverbil ka näiteks inglise keeles (Vanhove 2008: 348–349).

Kuulma võib eesti keeles esineda ka diskursuspartiklina, seda teise isiku ainsuses või mitmuses, käskivas kõneviisis (*kuule, kuulge*). Sellisel juhul on *kuulma* funktsioon tõmmata tähelepanu järgnevale kõnele. Samasugust kasutusmustrit on täheldatud ka inglise, sivu, hispaania ja itaalia keeles (San Roque jt 2018: 288–389). Eesti keeles võib diskursuspartiklina esineda ka kogemustajuverb *nägema* (*näed, näe, näete*). Tajuverbide partiklistumist on eesti keeles lähemalt uurinud Leelo Keevallik (2005, 2008).

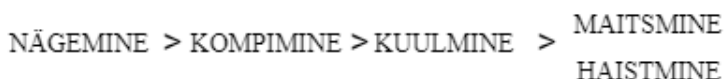
Samuti selgus III uurimusest, et eesti keele lõhna ja maitse kogemustaju põhi-verbid ei ole eriti produktiivsed tähenduse laienemise allikad. Selle poolset sarnaneb eesti keel enamike teiste Euroopa keeltega. Lisaks käituvad *maitset tundma* ja *lõhna tundma* ka korpuses teineteisega väga sarnaselt ning seega on neid ka väga keeruline lauseliste tunnuste põhjal eristada. Üks põhjusi selleks on nende verbide morfosüntaktiline koostis – tegemist on verbiühenditega, mille koosseisu kuulub partitiivne objekt. Seega ei saa kasutuses objekt enam varieeruda ning potentsiaalsete konstruktsioonide arv on seega piiratum. Verbide *maitset tundma* ja *lõhna tundma* eri tähendused ilmnevad eelkõige täiendi varieerumisel, nt *tunnen raha lõhna* vs. *tunnen supi lõhna*. Samas on ka selliste täiendite semantiline ulatus piiratud, nii *maitset tundma* kui ka *lõhna tundma* väljendavad mittefüüsilises tähenduses lühidat kontakti millegagi, millegi ootust või proovimist.

Maitset tundma ja *lõhna tundma* analüüs väitekirjas ühendab kahte eri tõlgendust ning on eelkõige sõltuv uurimisküsimusest, millele parajasti vastust otsitakse. Ühest küljest on *maitset tundma* ja *lõhna tundma* verbifraasid, kus verbi *tundma* objekt on vastavalt *maitse* või *lõhn*. Sellisel juhul väljendavad *maitset tundma* ja *lõhna tundma* verbi *tundma* kahte võimalikku tähendust ning on analüüsivad *tundma* semantilise varieerumisena. Samas on aga põhjendatud ka *maitset tundma* ja *lõhna tundma* analüüsimine eraldi leksikaalsete üksustena, seda eelkõige just põhitajuverbide paradigmast lähtudes (vt tabel 1). Paradigmast on lahtrite täitmisel

eelkõige oluline küsimus, kas keeles leidub selliseid vahendeid, millega väljendatakse ainult seda ühte tajumise modaalsust ja dünaamilist süsteemi. Kui küsime ”Kas eesti keeles leidub leksikaalseid vahendeid, mis väljendavad maitse tundmise ja lõhna tundmise kogemusi?”, siis on nendeks tõesti just *maitset tundma* ja *lõhna tundma*. Selles analüüsis ei ole tegemist enam lihtsalt verbifraasidega, vaid nad moodustavadki eraldi mõistelise ühendi.

Töö osade 1.3 ja 3.1–3.3 põhjal on võimalik koostada ka potentsiaalne eesti keele põhitajuverbide hierarhia. Laiapõhjalise tüpoloogilise valimi põhjal töötas hierarhia esimest korda välja Viberg (1984) ning esmalt oli see hierarhia mõeldud selleks, et kirjeldada ühest tajumodaalsusest teise toimuvate tähendusülekannete mustreid. Laiapõhjalise tüpoloogilise valimi põhjal koostas Viberg hierarhia, mille tipus (e kõige vasakul) on nägemistaju, sellele järgnevad vastavalt kuulmine ja kompimine ning võrdsel kohal kõige lõpus on maitmine ja haistmine (Viberg 1984: 147). Hierarhias kõrgemal (vasakul) asetsevat modaalsust väljendav verb võib laieneda tähenduselt väljendama ka sellest allpool olevaid modaalsusi, kuid mitte vastupidi. Tähendusülekanne võib modaalsusi ka vahele jätta. Lisaks on suurem tõenäosus, et kõrgemal positsioonil asetsevad verbid leksikaliseeritakse keeles lihtverbidega, nad varieeruvad rohkem ning on tihti keeles ka sagedasemad (Viberg 2001: 1297).

Eesti tajuverbide hierarhia on esitatud joonisel 1.



Joonis 1. Esialgne eesti keele tajuverbide hierarhia

Hierarhia tipus on eesti keele puhul samuti nägemine, nagu ka Vibergi algse hierarhias – nägemistaju põhiverbid on eesti keeles kõige sagedasemad (vt peatükk 1.3), samuti on verbil *nägema* palju tähendusi (I uurimus, peatükk 3.1) ning verbi morfosüntaktiline varieerumine on lai, st verb esineb paljudes erinevates morfosüntaktilistes kontekstides (III uurimus, peatükk 3.3). Teisel kohal aga asetseb eesti keeles kuulmise asemel kompimine. Eesti keeles on kompimistaju põhiverbid kaks korda sagedasemad kui kuulumisverbid, samuti selgus II ja III uurimusest, et kompimisverb *tundma* on võrreldes verbiga *kuulma* palju varieeruvam nii semantiliselt kui ka morfosüntaktilisel tasandil. *Tundma* tähendus laieneb hierarhias ka allapoole, väljendades lisaks *lõhna tundmist* ja *maitse tundmist*. Samuti on vasemal asetsevad verbid *nägema*, *tundma*, *kuulma* lihtverbid ning paremal asetsevad verbid *lõhna tundma* ja *maitset tundma* verbiühendid. Hierarhiat saab siiski pidada vaid esialgseks, kuna see põhineb suures osas ainult kogemustaju põhiverbidel. Kuigi tulevikus vajaks hierarhia uuesti analüüsimist ning kinnitamist, pakub see siiski eesti keele tajuverbide süsteemist huvitavat ülevaadet.

Tähenduse uurimise metodoloogia väärib arutelu eelkõige kognitiivse seman-

tika vaatepunktist. Lisaks sellele, et tähendust ei eksisteeri väljaspool keele kõneleja meelt, on kognitiivse teooria järgi tähenduse üks osa ka entsüklopeediline maailmateadmine (nt Fillmore 1985, Lakoff 1987, Talmy 2000). Maailmateadmine kätkeb endas kõike, mida on andnud meile meie maailmakogemus, sh teadmine nii meid ümbritsevast maailmast kui ka teadmised abstraktsetest mõistetest. Seega peaksime tähendust uurides esmalt küsima, kas ja kuidas me saame uurida kõike seda, millest tähendus koosneb, ja mis on meie tegelik kitsam uurimisobjekt iga uurimismeetodi puhul.

Väitekirjas kasutatud meetodid erinevadki üksteisest selle poolest, milliseid andmeid ja mille kohta need pakuvad. Korpusmeetodite puhul näiteks toetatakse põhimõttele, et keele uurimises on kõige olulisem arvestada keelekasutuse kontekstiga, kuna keele olemus seisneb eelkõige tema kasutuses. See tähendab, et kui koguda kokku suur hulk keele kasutamise kontekste, siis leiame sealt mustreid, mis võimaldavad meil luua teooriaid keele olemuse kohta (Glynn 2010, Glynn ja Fischer 2010). Seega kuigi me ei pääse korpuse toel vahetult ligi keelekasutaja meelele, saame siiski toetuda sellele, kuidas keelt on kasutatud.

Katselised meetodid pakuvad omakorda teistsuguseid andmeid, kuid ka katsete tulemused ei pruugi pakkuda otsust ligipääsu sellele, mis tegelikult keele kõneleja meeles on (vt ka Sandra ja Rice 1995, Cuyckens jt 1997, Gibbs ja Matlock 2001). Kõik väitekirjas kasutatud katsed näitavad küll seda, kuidas katseisik keelelisele stiimulile reageerib, kuid sellest ei saa siiski teha otsust järeldust selle keele kohta, kuidas keel tegelikult tema meeles esindatud ja organiseeritud on. Niisiis võib öelda, et tähendusele kui sellisele puudub otsene ligipääs nii korpuspõhistel meetoditel kui ka katselistel meetoditel. Samas aga on meil võimalik eri meetoditega saadud tulemusi võrrelda ja neid üksteise konteksti asetada. Seeläbi saame tükki tükki haaval koostada võimalikult terviklikku pilti sellest, mis tähendus on. Siin töötavad korpuspõhised ja katselised meetodid koos.

Lisaks erinevustele tõendite olemuses on katseliste ja korpuspõhiste meetodite puhul olulised kaalutlused ka valimi suurus või võimalus uurimisolukorda kontrollida (vt ka Gilquin ja Gries 2009, Arppe jt 2010). Näiteks on korpuspõhiste meetodite puhul uurijal eelis väga suure potentsiaalse valimi näol, tihti puudub aga teave selle kohta, millises kontekstis valimisse kuuluv keelematerjal on loodud. On aga ka spetsiifilisemaid eeliseid ja puudusi, mis puudutavad konkreetseid väitekirjas kasutatud meetodeid. Näiteks käitumisprofiili meetodi puhul on Glynn (2016) välja toonud, et tähenduste käsitlemine selle meetodi kohaselt ei peegelda kognitiivse semantika teoreetilisi seisukohti. Nimelt on polüseemia uurimisel käitumisprofiili meetodiga tarvis iga lause puhul ära määrata, mis tähendust verb konkreetse lauses kannab. See aga eeldab, et tähendust käsitletakse kui diskreetset üksust, st korruga saab sõnal olla ainult üks tähendus. Samas aga kognitiivne semantika käsitleb polüseemiat kui kategoriaalset üksust, kus erinevad tähendused võivad olla konkreetse kasutuses ka põimunud.

Lisaks eelnevalt välja toodud võimalikele probleemidele tähenduse korpuspõ-

hisel uurimisel, näitlikustab väitekirja III uurimus veel ühte sellega kaasnevat võimalikku probleemi. Mida polüseemsem on verb, seda rikkalikum on ka selle verbi morfosüntaktiline ja semantiline varieerumine ja vastupidi. Seega varieeruvad vähem polüseemsed verbid vähem ning nende varieerumist on lihtsam korpusvalimi põhjal ammendavalt kirjeldada. Samuti on vähem polüseemsete verbide tähendused diskreetsemad, näiteks *maitset tundma* eri tähendused on teineteisest lihtsamini eristatavad kui *nägema* eri tähendused. Väitekirja III uurimusest selgus seega, et vähem polüseemseid verbe *kuulma*, *maitset tundma* ja *lõhna tundma* ja nende kasutusmustreid on korpuses võimalik ammendavamalt kirjeldada kui väga polüseemsete verbide *nägema* ja *kuulma* kasutusmustreid.

Korpuspõhise lähenemise teel saadud tulemusi on mõistlik täiendada keelekatsetega (ja ka vastupidi). Katse võimaldab korpusandmetest vahetumalt uurida keelekasutaja arusaamu keelest. Ometigi on iga katseliigi -ja disaini puhul tarvilik arvestada selle omapäradega. Väitekirja I ja II uurimuses kasutati sorteerimiskatset: katseisiku ülesanne oli sorteerida elemendid rühmadesse etteantud tingimuse alusel. Väitekirja kaasatud uurimustes pidid katseisikud lauseid sorteerima lauses esineva tajuverbi tähenduse järgi. Siinkohal on aga oluline mõista, et stiimulite loomisega loob uurija ise taustsüsteemi, milles katseisik oma otsused langetab. Langacker (1987: 17) on öelnud, et keele kasutaja paigutab mingi elemendi mingisse rühma juhul, kui leidub mingisugune võimalik sarnasus selle elemendi ja rühma prototüüpse elemendi vahel. Näiteks II uurimuses sorteerisid katseisikud ühte rühma järgmised *tundma* tähendused: 'lõhna tundma', 'maitset tundma', 'valu tundma', 'temperatuuri tundma'. See aga ei tähenda, et keelekasutajate jaoks oleks need tähendused mõistelisel tasandil väga sarnased, pigem on sorteerimisotsuste taga see, et näiteks 'maitset tundma' ja 'valu tundma' on omavahel sarnasemad kui 'maitset tundma' ja 'emotsioonide tundmine' või 'maitset tundma' ja 'füüsilik tundma'. Seega peab nii korpuspõhiste kui ka katseliste meetodite rakendamisel olema teadlik võimalikest tingimustest ning meetodi olemuslikest piirangutest, mis võivad saadud tulemusi mõjutada.

Kokkuvõte

Väitekirjas esitatud tulemused näitavad, et eesti keele kogemustajuverbid on väga mitmekesised nii oma tähenduste kui ka kasutusmuurite poolest. Eesti keelele omased tajuverbide kasutusmuurid kinnitavad, et tajukogemuse mõistestamine oleneb keelest ning seda ümbritsevast kultuurist, teisest küljest peegeldavad eesti keele muurid ka kõiki maailma keeli ja inimesi ühendavaid kogemusi. Samuti näitlikustab väitekirja, kuidas erinevate meetodite kombineerimisega on võimalik lähemale pääseda ka niivõrd ligipääsmatule nähtusele, nagu seda on tähendus.

Väitekirja raames tehtu on avanud ka palju võimalusi edaspidiseks uurimistööks. Kindlasti väärib lähemat uurimist kompimistaju kogemusverb *tundma*, eriti selle teadmise ja aru saamisega seotud tähendused. Näiteks võiks neid tähen-

dusi uurida just eksplitsiitse ja implitsiitse teadmise seisukohast (Dienes ja Perner 1999, Ellis 2005, Jung ja Newen 2010, Haye ja Torres-Sahli 2017). Huvitavat uurimismaterjali seoses verbiga *tundma* leidub tulevikuks veelgi. Näiteks oleks omal kohal tüpoloogiline uurimus, mis keskenduks *tundma* kujunemisloole ning praegusele kasutusele Läänemere piirkonna keeltes. Samuti usun, et viljakas oleks uurida lähemalt tajuverbide kasutusi diskursuspartiklitena, seda eelkõige just võrdluses teiste keelte sarnaste partiklitega, kasutades näiteks Eesti suulise keele korpust (Rääbis 2013) ja Eesti keele spontaanse kõne foneetilist korpust (Lippus 2013). Tajuverbide partiklistumise uurimine just nende tajulise sisu poolest pakuks head täiendust Leelo Keevalliku (2005, 2008) töödele, kes on põhjalikult uurinud nende partiklite funktsioone ning arengut. Kuigi kirjutatud keelel põhinevad uurimused on Euroopa keelte puhul hetkel veel ülekaalus, keskendutakse üha enam ka suulise keele uurimisele. Diskursuspartiklite uurimine panustaks seega nii eesti keele uurimisse, aga maailma kontekstis ka laiemalt suulise keele uurimisse.

REFERENCES

- Arppe, Antti, Gaëtanelle Gilquin, 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.
- Baayen, R. Harald. 2014. Multivariate statistics. In Devyani Sharma and Robert J. Podesva, editors, *Research Methods in Linguistics*, Cambridge University Press, Cambridge, 337–372.
- Barlow, Michael and Suzanne Kemmer, editors. 2000. *Usage-Based Models of Language*. CSLI Publications, Stanford.
- BCE. 2014. Balanced Corpus of Estonian. <http://www.cl.ut.ee/korpused/grammatikakorpus/>.
- Beitel, Dinara A., Raymond W. Gibbs, Jr, and Paul Sanders. 2001. The Embodied approach to the Polysemy of the Spatial Preposition On. In Hubert Cuyckens and Britta Zawada, editors, *Polysemy in Cognitive Linguistics: Selected Papers from the Fifth International Cognitive Linguistics Conference*, John Benjamins Publishing Company, Amsterdam/Philadelphia, 241–260.
- Berez, Andrea L. and Stefan Th. Gries. 2008. In defense of corpus-based methods: A behavioral profile analysis of polysemous get in English. In Steven Moran, Darren S. Tanner, and Michael Scanlon, editors, *Proceedings of the 24th Northwest Linguistics Conference*, University of Washington Working Papers in Linguistics, Washington, Seattle, volume 27, 157–166.
- Born, Toomas. 1995. Eesti keele tajuverbid. Täendusallikad ja tähendusmuutused [Estonian perception verbs. Meaning sources and changes]. Tartu Ülikool.
- Bossomaier, Terry. 2012. The chemical senses. In *Introduction to the Senses: From Biology to Computer Science*, Cambridge University Press, Cambridge, 218–245. <https://doi.org/10.1017/CBO9781139016001>.
- Breiman, Leo. 2001. Random forests. *Machine Learning* 45(1): 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina, and R. Harald Baayen. 2007. Predicting the Dative Alternation. In Gerlof Bouma, Irene Krämer, and Joost Zwarts, editors, *Cognitive Foundations of Interpretation*, Royal Netherlands Academy of Science, Amsterdam, 69–94.
- Bresnan, Joan and Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86(1): 168–213.
- Clark, Herbert H. 1973. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior* 12(4): 335–359. [https://doi.org/10.1016/S0022-5371\(73\)80014-3](https://doi.org/10.1016/S0022-5371(73)80014-3).
- Classen, C. 2005. *The Book of Touch*. Sensory Formations. Bloomsbury Academic.

- Classen, Constance. 1993. *Worlds of Sense: Exploring the Senses in History and across Cultures*. Routledge, London ; New York.
- Croft, William and Alan D. Cruse. 2004. *Cognitive Linguistics*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511803864>.
- Croijmans, Ilja and Asifa Majid. 2016. Not All Flavor Expertise Is Equal: The Language of Wine and Coffee Experts. *PLOS ONE* 11(6): e0155845. <https://doi.org/10.1371/journal.pone.0155845>.
- Cuyckens, Hubert, Dominiek Sandra, and Sally Rice. 1997. Towards an empirical lexical semantics. In Birgit Smieja, editor, *Human Contact through Language and Linguistics*, Lang, Frankfurt am Main Berlin Bern New York Paris Wien, number 31 in Duisburger Arbeiten Zur Sprach- Und Kulturwissenschaft, 35–54.
- Dienes, Zoltan and Josef Perner. 1999. A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences* 22(5): 735–808.
- Divjak, Dagmar. 2015. Exploring the grammar of perception. A case study using data from Russian. *Functions of Language* 22(1): 44–68.
- Divjak, Dagmar and Nick Fieller. 2014. Cluster analysis. Finding structure in linguistic data. In Dylan Glynn and Justyna A. Robinson, editors, *Corpus Methods for Semantics. Quantitative Studies in Polysemy and Synonymy*, John Benjamins Publishing Company, Amsterdam/Philadelphia, 405–441.
- Ellis, Rod. 2005. Measuring implicit and explicit knowledge of a second language: A Psychometric Study. *Studies in Second Language Acquisition* 27(02): 141–172. <https://doi.org/10.1017/S0272263105050096>.
- Enghels, Renata and Marlies Jansegers. 2013. On the crosslinguistic equivalence of sentir(e) in Romance languages: A contrastive study in semantics. *Linguistics* 51(5): 957–991.
- Erelt, Mati, editor. 2003. *Estonian Language*, volume 1 of *Linguistica Uralica. Supplementary Series*. Estonian Academy Publishers, Tallinn.
- Erelt, Mati and Helle Metslang, editors. 2017. *Eesti Keele Süntaks [Estonian Syntax]*. Number III in *Eesti Keele Varamu*. Tartu Ülikool kirjastus, Tartu.
- etTenTen. 2013. Estonian Web Corpus. <https://keeleeveeb.ee/dict/corpus/etten-ten/>.
- Evans, Nicholas and David Wilkins. 2000. In the Mind's Ear: The Semantic Extensions of Perception Verbs in Australian Languages. *Language: Journal of the Linguistic Society of America* 76(3): 546–592.
- Fillmore, C.J. 1985. Frames and the Semantics of Understanding. *Quaderni di Semantica* 6(2): 222–254.
- Firth, John R. 1957. *Papers in Linguistics*. Oxford University Press, London.
- Franco, Karlien, Dirk Geeraerts, Dirk Speelman, and Roeland Van Hout. 2019. Concept characteristics and variation in lexical diversity in two Dutch dialect areas. *Cognitive Linguistics* 30(1): 205–242. <https://doi.org/10.1515/cog-2017-0136>.

- Geeraerts, Dirk. 2006. Chapter 5. Vagueness's puzzles, polysemy's vagaries. In René Dirven, John R. Taylor, Ronald W. Langacker, and Dirk Geeraerts, editors, *Words and Other Wonders*, Mouton de Gruyter, Berlin, New York, volume 33, 99–148.
- Gibbs, Jr, Raymond W. 2006. *Embodiment and Cognitive Science*. Cambridge University Press, Cambridge.
- Gibbs, Jr, Raymond W. and Teenie Matlock. 2001. Psycholinguistic Perspectives on Polysemy. In Hubert Cuyckens and Britta Zawada, editors, *Polysemy in Cognitive Linguistics: Selected Papers from the Fifth International Cognitive Linguistics Conference*, John Benjamins Publishing Company, Amsterdam/Philadelphia, 213–239.
- 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.
- Glynn, Dylan. 2010. Corpus-driven Semantics. In Dylan Glynn and Kerstin Fischer, editors, *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, Walter de Gruyter GmbH&Co. KG, Berlin/New York, number 46 in *Cognitive Linguistics Research*, 1–43.
- Glynn, Dylan. 2014. The many uses of run. Corpus methods and Socio-Cognitive Semantics. In Dylan Glynn and Justyna A. Robinson, editors, *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*, John Benjamins Publishing Company, number 43 in *Human Cognitive Processing (HCP). Cognitive Foundations of Language Structure and Use*, 117–144.
- Glynn, Dylan. 2016. Quantifying polysemy: Corpus methodology for prototype theory. *Folia Linguistica* 50(2): 413–447.
- Glynn, Dylan and Kerstin Fischer, editors. 2010. *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*. De Gruyter Mouton, Berlin, New York. <https://doi.org/10.1515/9783110226423>.
- Glynn, Dylan and Justyna A. Robinson, editors. 2014. *Corpus Methods for Semantics*, volume 43 of *Human Cognitive Processing (HCP). Cognitive Foundations of Language Structure and Use*. John Benjamins Publishing Company, Amsterdam/Philadelphia.
- Grafmiller, Jason and Benedikt Szmrecsanyi. 2018. Mapping out particle placement in Englishes around the world: A study in comparative sociolinguistic analysis. *Language Variation and Change* 30(3): 385–412. <https://doi.org/10.1017/S0954394518000170>.
- Gries, Stefan Th. 2006. Corpus-based methods and cognitive semantics: The many senses of to run. In Stefan Th. Gries and Anatol Stefanowitsch, editors, *Corpora in Cognitive Linguistics: Corpus-Based Approaches to Syntax and Lexis*, Mouton de Gruyter, Berlin/New York, 57–99.
- 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. 2013. *Statistics for Linguistics with R: A Practical Introduction*. De Gruyter, Berlin, Boston. <https://doi.org/10.1515/9783110307474>.
- Gries, Stefan Th. and Dagmar Divjak. 2009. Behavioral profiles: A corpus-based approach to cognitive semantic analysis. In Vyvyan Evans and Stéphanie Pourcel, editors, *New Directions in Cognitive Linguistics*, John Benjamins Publishing Company, Amsterdam/Philadelphia, 57–77.
- Gries, Stefan Th. and Dagmar Divjak. 2010. Quantitative approaches in usage-based cognitive semantics. In Dylan Glynn and Kerstin Fischer, editors, *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, De Gruyter Mouton, number 46 in Cognitive Linguistics Research, 331–354.
- Harris, Zellig S. 1954. Distributional Structure. *Word* 10(2-3): 146–162. <https://doi.org/10.1080/00437956.1954.11659520>.
- Haye, Andrés and Manuel Torres-Sahli. 2017. To feel is to know relations: James' concept of stream of thought and contemporary studies on procedural knowledge. *New Ideas in Psychology* 46: 46–55. <https://doi.org/10.1016/j.newideaspsych.2017.02.001>.
- Hollman, Liivi and Urmas Sutrop. 2010. Basic Color Terms in Estonian Sign Language. *Sign Language Studies* 11(2): 130–157. <https://doi.org/10.1353/sls.2010.0011>.
- Howes, David, editor. 1991. *The Varieties of Sensory Experience: A Sourcebook in the Anthropology of the Senses*. Anthropological Horizons. University of Toronto Press, Toronto ; Buffalo.
- Howes, David, editor. 2005. *Empire of the Senses: The Sensual Culture Reader*. Sensory Formations Series. Berg, Oxford; New York.
- Huumo, Tuomas. 2010. Is perception a directional relationship? On directionality and its motivation in Finnish expressions of sensory perception. *Linguistics* 48(1). <https://doi.org/10.1515/ling.2010.002>.
- Ibarretxe-Antuñano, Iraide. 2006. Cross-Linguistic Polysemy in Tactile Verbs. In *Cognitive Linguistics Investigations across Languages, Fields, and Philosophical Boundaries*, John Benjamins Publishing Company, Amsterdam/Philadelphia, number 15 in Human Cognitive Processing (HCP). Cognitive Foundations of Language Structure and Use, 235–253.
- Ibarretxe-Antuñano, Iraide. 2008. Vision Metaphors for the Intellect: Are They Really Cross-Linguistic? *Atlantis. Journal of the Association of Anglo-American Studies* 30(1): 15–33.
- Janda, Laura A. and Valery D. Solovyev. 2009. What constructional profiles reveal about synonymy: A case study of Russian words for sadness and happiness. *Cognitive Linguistics* 20(2): 295–324. <https://doi.org/10.1515/COGL.2009.018>.
- Jansegers, Marlies and Stefan Th. Gries. 2020. Towards a dynamic behavioral profile: A diachronic study of polysemous sentir in Spanish. *Corpus Linguistics and Linguistic Theory* 16(1): 145–187. <https://doi.org/10.1515/cllt-2016-0080>.

- Jansegers, Marlies, Clara Vanderschueren, and Renata Enghels. 2015. The polysemy of the Spanish verb sentir: A behavioral profile analysis. *Cognitive Linguistics* 26(3): 381–421.
- Johnson, Mark. 1987. *The Body in the Mind. The Bodily Basis of Meaning, Imagination, and Reason*. The University of Chicago Press, Chicago.
- Jung, Eva-Maria and Albert Newen. 2010. Knowledge and abilities: The need for a new understanding of knowing-how. *Phenomenology and the Cognitive Sciences* 9(1): 113–131. <https://doi.org/10.1007/s11097-009-9129-3>.
- Keevallik, Leelo. 2005. Suhtluskeele uurimine partiklistunud verbivormide näitel [Investigating spoken language: Evidence from verb forms turned into particles]. *Keel ja Kirjandus* 7, 8: 535–548, 630–642.
- Keevallik, Leelo. 2008. Internal development and borrowing of pragmatic particles: Estonian "vaata/vat" 'look', "nääd" 'you see', and "vot". *Finnisch-Ugrische Mitteilungen* 30/31: 23–54.
- Klavan, Jane. 2017. Pitting corpus-based classification models against each other: A case study for predicting constructional choice in written Estonian. *Corpus Linguistics and Linguistic Theory* 16(2). <https://doi.org/10.1515/cllt-2016-0010>.
- Labi, Kanni. 2006. Eesti regilaulude verbisemantika [Verb semantics of Estonian runic songs]. Tartu Ülikool.
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. The University of Chicago Press, Chicago and London.
- Lakoff, George and Mark Johnson. 1980. *Metaphors We Live By*. The University of Chicago Press, Chicago and London.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Volume 1. Theoretical Prerequisites*. Stanford University Press, Stanford, California.
- Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar. Volume II. Descriptive Application*. Stanford University Press, Stanford, Calif.
- Langemets, Margit, Mai Tiits, Tiia Valdre, Leide Veskis, Ülle Viks, and Piret Voll, editors. 2009. *Eesti Keele Seletav Sõnaraamat [Explanatory Dictionary of Estonian]*. Eesti Keele Sihtasutus, Tallinn.
- Lenci, Alessandro. 2008. Distributional semantics in linguistic and cognitive research. *Italian Journal of Linguistics* 20: 1–31.
- Levshina, Natalia. 2015. *How to Do Linguistics with R. Data Exploration and Statistical Analysis*. John Benjamins Publishing Company, Amsterdam/Philadelphia.
- Lippus, Pärtel. 2013. Phonetic Corpus of Estonian Spontaneous Speech v.1.0.0 <https://doi.org/10.15155/1-00-0000-0000-0000-00074L>.
- Majid, Asifa and Niclas Burenhult. 2014. Odors are expressible in language, as long as you speak the right language. *Cognition* 130(2): 266–270.
- Majid, Asifa and Nicole Kruspe. 2018. Hunter-Gatherer Olfaction Is Special. *Current Biology* 28(3): 409–413.e2. <https://doi.org/10.1016/j.cub.2017.12.014>.

- Majid, Asifa and Stephen C. Levinson. 2011. The Senses in Language and Culture. *The Senses and Society* 6(1): 5–18. <https://doi.org/10.2752/174589311X12893982233551>.
- Majid, Asifa, Seán G. Roberts, Ludy Cilissen, Karen Emmorey, Brenda Nicodemus, Lucinda O’Grady, Bencie Woll, Barbara LeLan, Hilário de Sousa, Brian L. Cansler, Shakila Shayan, Connie de Vos, Gunter Senft, N. J. Enfield, Rogayah A. Razak, Sebastian Fedden, Sylvia Tufvesson, Mark Dingemanse, Ozge Ozturk, Penelope Brown, Clair Hill, Olivier Le Guen, Vincent Hirtzel, Rik van Gijn, Mark A. Sicoli, and Stephen C. Levinson. 2018. Differential coding of perception in the world’s languages. *Proceedings of the National Academy of Sciences* 115(45): 11369–11376. <https://doi.org/10.1073/pnas.1720419115>.
- Metsmägi, Iris, Meeli Sedrik, and Sven-Erik Soosaar, editors. 2012. *Eesti Etümooloogiasõnaraamat [Etymological Dictionary of Estonian]*. Eesti Keele Sihtasutus, Tallinn.
- Oja, Vilja. 2001. *Linguistic Studies of Estonian Colour Terminology*. Dissertationes Philologiae Estonicae Universitatis Tartuensis 18, University of Tartu, Tartu.
- Oja, Vilja. 2014. The motivational analysis of some Finnic colour terms. In Wendy Anderson, Carole P. Biggam, Carole Hough, and Christian Kay, editors, *Colour Studies*, John Benjamins Publishing Company, Amsterdam, 93–108. <https://doi.org/10.1075/z.191.06oja>.
- Oja, Vilja. 2016. Moodustusviisi ja tähenduse vahekorra eesti värvinimetustes [Estonian colour names: Relations between structure and meaning]. *Mäetagused* 64: 99–114. <https://doi.org/10.7592/MT2016.64.oja>.
- Oja, Vilja and Mari Uusküla. 2010. Mõnest värvinimetusest ja nende tähendusvahekordadest eesti ja soome keeles [On some colour terms and their semantic relationships in Estonian and Finnish]. *Eesti Rakenduslingvistika Ühingu aastaraamat. Estonian Papers in Applied Linguistics* 6: 195–205.
- O’Meara, Carolyn and Asifa Majid. 2020. Anger stinks in Seri: Olfactory metaphor in a lesser-described language. *Cognitive Linguistics* 0(0). <https://doi.org/10.1515/cog-2017-0100>.
- Peirsman, Yves, Dirk Geeraerts, and Dirk Speelman. 2010. The automatic identification of lexical variation between language varieties. *Natural Language Engineering* 16(4): 469–491. <https://doi.org/10.1017/S1351324910000161>.
- Proos, Mariann. 2019a. Data & script for the article ”Polysemy of the Estonian perception verb ’nägema’ ’to see’” <https://doi.org/10.17605/OSF.IO/GM5SE>.
- Proos, Mariann. 2019b. Polysemy of the Estonian perception verb ”nägema” ’to see’. In Laura J. Speed, Carolyn O’Meara, Lila San Roque, and Asifa Majid, editors, *Perception Metaphors*, John Benjamins Publishing Company, number 19 in *Converging Evidence in Language and Communication Research (CELCR)*, 231–252.

- Proos, Mariann. 2020a. Data & script for the article "Feeling your neighbour: An experimental approach to the polysemy of *tundma* 'to feel' in Estonian". *Open Science Framework* .
- Proos, Mariann. 2020b. Data & script for the article "Usage patterns of Estonian experience perception verbs: A random forest approach". *Open Science Framework* <https://doi.org/10.17605/OSF.IO/JSHVF>.
- Proos, Mariann. 2020c. Feeling your neighbour: An experimental approach to the polysemy of *tundma* 'to feel' in Estonian. *Language and Cognition* 12(2): 282–309. <https://doi.org/10.1017/langcog.2019.44>.
- Proos, Mariann. 2020d. Usage patterns of Estonian experience perception verbs: A random forest approach. *SKY Journal of Linguistics* 33: 2020: 67–101.
- Rääbis, Andriela. 2013. Corpus of Spoken Estonian <https://doi.org/10.15155/1-00-0000-0000-0000-00077L>.
- Raukko, Jarno. 2003. Polysemy as flexible meaning: Experiments with English *get* and Finnish *pitää*. In Brigitte Nerlich, Zazie Todd, Vimala Herman, and David D. Clarke, editors, *Polysemy*, De Gruyter Mouton, Berlin, New York, 161–194. <https://doi.org/10.1515/9783110895698.161>.
- Rosch, Eleanor. 1978. Principles of Categorization. In *Cognition and Categorization*, Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey, 27–48.
- Rosch, Eleanor and Carolyn B. Mervis. 1975. Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology* 7: 573–605.
- San Roque, Lila, Kobin H. Kendrick, Elisabeth Norcliffe, Penelope Brown, Rebecca Defina, Mark Dingemanse, Tyko Dirksmeyer, N. J. Enfield, Simeon Floyd, Jeremy Hammond, Giovanni Rossi, Sylvia Tufvesson, Saskia van Putten, and Asifa Majid. 2014. Vision verbs dominate in conversation across cultures, but the ranking of non-visual verbs varies. *Cognitive Linguistics* 26(1): 31–60. <https://doi.org/10.1515/cog-2014-0089>.
- San Roque, Lila, Kobin H. Kendrick, Elisabeth Norcliffe, and Asifa Majid. 2018. Universal meaning extensions of perception verbs are grounded in interaction. *Cognitive Linguistics* 29(3): 371–406. <https://doi.org/10.1515/cog-2017-0034>.
- 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.
- Schmid, Hans-Jörg. 2010. Does frequency in text instigate entrenchment in the cognitive system? In Dylan Glynn and Kerstin Fischer, editors, *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, Mouton de Gruyter, Berlin/New York, number 46 in Cognitive Linguistics Research, 101–137.
- Sepper, Maria-Maren. 2006. *Indirektaal Eesti 19. Sajandi Lõpu Ja 20. Sajandi Aja- Ja Ilukirjanduskeeles*. Magistritöö, Tallinna Ülikool, Tallinn.
- Sjöström, Sören. 1999. From Vision to Cognition. A study of Metaphor and Polysemy in Swedish. In Jens Allwood and Peter Gärdenfors, editors, *Cognitive Se-*

- mantics. Meaning and Cognition*, John Benjamins Publishing Company, Amsterdam/Philadelphia, 67–85.
- Speed, Laura J., Carolyn O’Meara, Lila San Roque, and Asifa Majid, editors. 2019. *Perception Metaphors*. Number volume 19 in *Converging Evidence in Language and Communication Research (CELCR)*. John Benjamins Publishing Company, Amsterdam ; Philadelphia.
- Stefanowitsch, Anatol. 2010. Empirical Cognitive Semantics. In Dylan Glynn and Kerstin Fischer, editors, *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, De Gruyter Mouton, Berlin/New York, number 46 in *Cognitive Linguistics Research*, 355–380.
- Storch, Anne. 2013. Knowing, Smelling and Telling Tales in Luwo. In Anne Storch and Alexandra Aikhenvald, editors, *Perception and Cognition in Language and Culture*, Brill, 47–68. https://doi.org/10.1163/9789004210127_03.
- Storch, Anne and Alexandra Aikhenvald, editors. 2013a. *Perception and Cognition in Language and Culture*. Brill. <https://doi.org/10.1163/9789004210127>.
- Storch, Anne and Alexandra Y. Aikhenvald. 2013b. Linguistic Expression of Perception and Cognition: A Typological Glimpse. In Anne Storch and Alexandra Aikhenvald, editors, *Perception and Cognition in Language and Culture*, Brill, 1–45. https://doi.org/10.1163/9789004210127_02.
- Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9(1): 307.
- Strobl, Carolin, James Malley, and Gerhard Tutz. 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14(4): 323–348. <https://doi.org/10.1037/a0016973>.
- Sutrop, Urmas. 1998. Eestlase maitse sõnad [Taste words in Estonian]. *Keel ja Kirjandus* 7: 446–459.
- Sutrop, Urmas. 2001. Lõhnad meie keeles [Smells in our language]. *Keel ja Kirjandus* 1: 5–13.
- Sutrop, Urmas. 2002. *The Vocabulary of Sense Perception in Estonian: Structure and History*. Number VII in *Opuscula Fenno-Ugrica Gottingensia*. Peter Lang, Frankfurt am Main.
- Sweetser, Eve E. 1990. *From Etymology to Pragmatics. Metaphorical and Cultural Aspects of Semantic Structure*. Number 54 in *Cambridge Studies in Linguistics*. Cambridge University Press, Cambridge.
- Szmrecsanyi, Benedikt, Jason Grafmiller, and Laura Rosseel. 2019. Variation-Based Distance and Similarity Modeling: A Case Study in World Englishes. *Frontiers in Artificial Intelligence* 2: 23. <https://doi.org/10.3389/frai.2019.00023>.
- Tagliamonte, Sali A. and R. Harald Baayen. 2012. Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice.

- Language Variation and Change* 24(2): 135–178. <https://doi.org/10.1017/S0954394512000129>.
- Talmy, Leonard. 2000. *Toward a Cognitive Semantics. Volume I: Concept Structuring Systems*. The MIT Press, Cambridge, London.
- Talmy, Leonard. 2003. *Toward a Cognitive Semantics. Volume II: Typology and Process in Concept Structuring*. MIT Press, Cambridge, Mass.
- Team, R Core. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing <https://www.r-project.org/>.
- Troche, Joshua, Sebastian Crutch, and Jamie Reilly. 2014. Clustering, hierarchical organization, and the topography of abstract and concrete nouns. *Frontiers in Psychology* 5: 1–10. <https://doi.org/10.3389/fpsyg.2014.00360>.
- Troche, Joshua, Sebastian J. Crutch, and Jamie Reilly. 2017. Defining a Conceptual Topography of Word Concreteness: Clustering Properties of Emotion, Sensation, and Magnitude among 750 English Words. *Frontiers in Psychology* 8: 1–15. <https://doi.org/10.3389/fpsyg.2017.01787>.
- Tummers, Jose, Kris Heylen, and Dirk Geeraerts. 2005. Usage-based approaches in Cognitive Linguistics: A technical state of the art. *Corpus Linguistics and Linguistic Theory* 1(2). <https://doi.org/10.1515/cllt.2005.1.2.225>.
- Usoniene, Aurelia. 2001. On direct/indirect perception with verbs of seeing and seeming in English and Lithuanian. *Lund University, Dept. of Linguistics. Working Papers* 48: 163–182.
- Uusküla, Mari. 2008. *Basic colour terms in Finno-Ugric and Slavonic languages: Myths and facts*. Number 9 in *Dissertationes linguisticae Universitatis Tartuensis*. Tartu Ülikooli Kirjastus, Tartu.
- Uusküla, Mari and David L. Bimler. 2020. When does “bright” mean “prototypical”? Color-term modifiers in eight European languages, examined with color-survey data. *Journal of the Optical Society of America A* 37(5): A305. <https://doi.org/10.1364/JOSAA.382399>.
- Uusküla, Mari, Liivi Hollman, and Urmas Sutrop. 2012. Basic colour terms in five Finno-Ugric languages and Estonian Sign Language: A comparative study. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 3(1): 47–86. <https://doi.org/10.12697/jeful.2012.3.1.02>.
- Uusküla, Mari and Urmas Sutrop. 2014. Põhivärvinimed läänemere ümbruse keeltes: Areaalsed, universaalsed või relativistlikud [Basic colour terms around the Baltic sea: Areal, universal, or relativistic?]. *Keel ja Kirjandus* 8-9: 670–683.
- van Putten, Saskia. 2020. Perception verbs and the conceptualization of the senses: The case of Avatime. *Linguistics* 58(2): 425–462. <https://doi.org/10.1515/ling-2020-0039>.
- Vanhove, Martine. 2008. Semantic associations between sensory modalities, prehension and mental perceptions. A crosslinguistic perspective. In Martine Vanhove, editor, *From Polysemy to Semantic Change. Towards a Typology*

- of *Lexical Semantic Associations*, John Benjamins Publishing Company, Amsterdam/Philadelphia, number 106 in Studies in Language Companion Series, 341–370.
- Viberg, Åke. 1984. The verbs of perception: A typological study. *Linguistics* 21: 123–162.
- Viberg, Åke. 2001. Verbs of perception. In H Martin, E König, W Oesterricher, and W Raible, editors, *Language Typology and Language Universals: An International Handbook*, Walter de Gruyter, Berlin and New York, 1294–1309.
- Viberg, Åke. 2005. The lexical typological profile of Swedish mental verbs. *Languages in Contrast* 5(1): 121–157. <https://doi.org/10.1075/lic.5.1.09vib>.
- Viberg, Åke. 2015. Sensation, perception and cognition: Swedish in a typological-contrastive perspective. *Functions of Language* 22(1): 96–131. <https://doi.org/10.1075/fo1.22.1.05vib>.
- Winter, Bodo. 2013. Linear models and linear mixed effects models in R with linguistic applications. *arXiv:1308.5499 [cs]* .
- Winter, Bodo, Marcus Perlman, and Asifa Majid. 2018. Vision dominates in perceptual language: English sensory vocabulary is optimized for usage. *Cognition* 179: 213–220. <https://doi.org/10.1016/j.cognition.2018.05.008>.
- Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. Blackwell Publishers Ltd., Oxford.
- Wnuk, Ewelina and Asifa Majid. 2014. Revisiting the limits of language: The odor lexicon of Maniq. *Cognition* 131(1): 125–138. <https://doi.org/10.1016/j.cognition.2013.12.008>.

PUBLICATIONS

CURRICULUM VITAE

Name: Mariann Proos
Date of birth: 26.09.1991
Citizenship: Estonian
Address: Institute of Estonian and General Linguistics,
University of Tartu,
Jakobi 2-446, Tartu
E-mail: mariann.proos@ut.ee

Education:
2016–2020 University of Tartu, PhD student (Estonian and Finno-Ugric linguistics)
01.09.2017–
30.08.2018 Radboud University Nijmegen. CLS - Centre for Language Studies. Visiting researcher.
2014–2016 University of Tartu, MA (Estonian and Finno-Ugric linguistics), *cum laude*
01.09.2015–
01.02.2016 KU Leuven. Advanced Studies in Linguistics. Semester abroad.
2011–2014 University of Tartu, BA (Estonian and Finno-Ugric linguistics)
1998–2010 Kadrioru Saksa Gümnaasium

Professional experience:
2020–... University of Tartu, Faculty of Arts and Humanities, Dean's Office, event coordinator for the Graduate School of Arts and Humanities
2019–... University of Tartu, Faculty of Arts and Humanities, Institute of Estonian and General Linguistics, Junior Research Fellow in Estonian and Finno-Ugric Linguistics

ELULOOKIRJELDUS

Nimi: Mariann Proos
Sünnikuupäev: 26.09.1991
Kodakondsus: Eesti
Aadress: Eesti ja üldkeeleteaduse instituut,
Tartu Ülikool,
Jakobi 2-446, Tartu
E-mail: mariann.proos@ut.ee

Haridus:
2016–2020 Tartu Ülikool, Eesti ja soome-ugri keeleteadus,
doktoriõpe
01.09.2017–
30.08.2018 Radboud University Nijmegen. CLS - Centre for
Language Studies. Külalisdoktorant
2014–2016 Tartu Ülikool, Eesti ja soome-ugri keeleteadus,
magistrikraad (*cum laude*)
01.09.2015–
01.02.2016 KU Leuven. Advanced Studies in Linguistics.
Erasmuse välissemester.
2011–2014 Tartu Ülikool, Eesti ja soome-ugri keeleteadus,
bakalaureusekraad
1998–2010 Kadrioru Saksa Gümnaasium, keskharidus

Töökogemus:
2020–... Tartu Ülikool, Humanitaarteaduste ja kunstide
valdkonna dekanaat, doktorikoolide ürituste
koordinaator
2019–... Tartu Ülikool, Humanitaarteaduste ja kunstide
valdkond, eesti ja üldkeeleteaduse instituut, eesti ja
soome-ugri keeleteaduse nooremteadur

DISSERTATIONES LINGUISTICAE UNIVERSITATIS TARTUENSIS

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

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