



# Oblique complements in Estonian: A corpus perspective

MARI AIGRO 

*University of Tartu*

VIRVE-ANNELI VIHMAN 

*University of Tartu*

(Received 26 April 2021; revised 29 November 2022)

This study focuses on Estonian verb-complement structures, which include oblique (non-canonically marked) complements marked in spatial cases. Not all approaches agree on whether canonical arguments and oblique complements have argument status of the same type, but they do mostly agree that the two types of complement markings are used by different types of verbs. First, oblique case is viewed as always indexing the original semantics of the case (direct semantics), that is OSUTAMA ‘point at’ selecting an allative (‘onto’) complement. Second, oblique case usage is seen as referring to a restricted set of syntactic relations (indirect semantics), that is Estonian allative and adessive being used for marking Experiencers. In any case, oblique complement verbs are viewed as more semantically restricted than canonical object verbs. This study tests these two hypotheses in a quantitative corpus approach. In a non-semantically extracted sample of verbs (n = 232), it compares the lexical-semantic transitivity of oblique and canonical complement verbs in order to investigate the degree to which indirect semantic effects differentiate between the two types of verbs. In addition, it outlines direct semantic effects between oblique case frames in terms of semantic roles. Finally, it investigates the way these patterns are related to the cases’ individual grammaticalisation degrees.

KEYWORDS: argument status, Estonian, non-canonical arguments, spatial case, transitivity

## 1. INTRODUCTION

This study focuses on the Estonian verb-complement structures in which complements are marked with spatial cases, as in Example (1):

- (1) pakkumine ei sõltu ainult hinnast<sup>1</sup>  
 offer NEG depend.CNG only price:ELA  
 ‘The offer does not only depend on the price’.  
 (Balanced Corpus of Estonian, henceforth, BCE)

[1] Abbreviations in glosses follow the Leipzig Glossing Rules.

Oblique complements constitute an area of interest for a variety of reasons. On one hand, their argument status has been subject to debate. Traditional approaches view case as an inherently bifurcated system, in which some cases ('direct' or 'syntactic') have the ability to mark arguments, and others ('oblique' or 'semantic') mark adjuncts. For this reason, oblique complements are categorised by some researchers as a subclass of adverbials, as is done in most Estonian reference grammars and approaches (Kettunen 1924, Erelt 2003, Veismann et al. 2017), by others as lying somewhere between arguments and adverbials (Blake 2001, Butt 2006) and by others still as regular arguments no different from canonically case-marked direct objects (Nichols 1983, 1984).

However, most researchers agree on one aspect of oblique complements: that verbs which mark complements non-canonically tend to differ from verbs with canonical complements in terms of their semantics. These semantic distinctions can be broadly seen as belonging to two distinct types, coined direct and indirect semanticity by Nichols (1983: 186), and referred to as direct/indirect semantics or semantic effects in the present paper. Direct semantic effects occur when the selected case directly indexes the semantic role of the complement or the lexical semantics of the verb. For instance, the Estonian verb *INVESTTEERIMA* 'invest in' takes complements in illative ('into'): spatial meaning is entailed in the semantics of both the verb and the complement. However, with indirect semantic effects, case usage refers to a syntactic relation which is restricted in a way that is not directly related to the original meaning of the case. Examples include Estonian Experiencers marked in spatial cases, such as dative-like functions of allative and adessive case.

Indirect semantics has received relatively more attention during the past decades than direct semantics. Topical research includes approaches which focus on broad cross-linguistic patterns, outlining lexical, morphological and syntactic factors attracting non-canonical case marking. Regarding lexical semantics, verbs taking oblique complements tend to be associated with some type of decreased transitivity, often accompanied by increased stativity (Tsunoda 1981, Blake 2001, Malchukov 2005, de Hoop & de Swart 2009, Lestrade 2010). However, many of these approaches are solely focused on the non-canonically marked complements of verbs, linking their choice of case to factors broadly outlined by the transitivity parameters in Hopper & Thompson (1980), also figuring in Tsunoda (1981), Dowty (1991). It is less common to investigate which indirect semantic effects constitute fundamental distinctions between oblique and canonical complement verbs in a single language, focusing on verbs where the case on complements does not alternate.

Direct semantic effects have received much less attention in the context of oblique complement marking. While both types of semantic effects are regarded as reasons for categorically differentiating between the argument status of oblique and canonical complements (Blake 2001), direct semantic effects constitute the more profound evidence of the two, because they indicate that even as a complement marker, oblique case is still used with some of its 'semantic' sense and has not acquired an entirely bleached, abstract syntactic function. However, there are no

quantitative studies showing that this effect is a general feature of oblique complement verbs. Furthermore, direct semantics appears to be strongly linked to the grammaticalisation of cases. In languages where cases can mark both arguments and adjuncts, semantic cases are expected to vary in terms of the degree to which they are multifunctional (Nichols 1983, 1984). As semantic case markers tend to grammaticalise into syntactic case markers (Bybee 1985, Hopper & Traugott 2003, Lestrade 2010), direct semantic effects are expected to be found more often with the less grammaticalised cases than the more grammaticalised cases.

This paper aims to investigate the two types of semantic effects in Estonian oblique complement verbs. Estonian constitutes a rich resource for investigating oblique complements. Out of the 14 morphological cases of Estonian, three are usually regarded as canonical argument-marking cases: partitive, genitive and nominative. The rest are regarded as semantic cases, seven of which have been noted to function as markers of oblique, non-canonical complements (six spatial cases and comitative, Veismann et al. 2017: 319). These spatial cases include allative ('onto'), adessive ('on'), ablative ('from on top of'), illative ('into'), inessive ('in') and elative ('from in') (Veismann et al. 2017).

This study has two aims. First, it takes a multilevel approach to stativity, decomposed into more basic notions, to explore the indirect semantic effects in Estonian verbs taking complements marked in spatial cases. Instead of merely analysing the way stativity is expressed in such verbs, the investigation is based on a comparison with a control group composed of verbs taking canonically marked complements. This allows us to test the hypothesis that the two groups of verbs differ in terms of lexical stativity (transitivity) and propose further hypotheses for cross-linguistic research, regarding the motivation underlying a language synchronically using a wide range of complement-marking strategies.

Second, the paper tackles direct semantic effects by outlining the semantic roles associated with Estonian oblique complement verbs. It also investigates a link between case-based semantic role patterns and the individual degree of grammaticalisation of these cases, as measured through productivity. The paper offers a more comprehensive picture of oblique complements. Hence, on the one hand the paper addresses the explanatory perspective on the phenomenon, and on the other hand it maps the degree to which each spatial case is bleached in the system of oblique complements, linking them to adjunct and canonical complement markers in a meaningful way.

In this study, the term 'oblique complement verbs', sometimes known as verbs with non-canonical complements, denotes specifically the verbs which select elements marked in spatial cases. Most of these elements are referred to as semantic arguments (event participants) in approaches contrasting a slightly wider range of semantic arguments with a narrow range of syntactic arguments (for example, Ackema 2015). Physical locations are excluded, though the participanthood and argumenthood of locations is an interesting topic for another study. Verbs taking complements marked in partitive, genitive and/or nominative are regarded as canonical complement verbs.

Although verbs with a particular type of complement structure constitute the subject matter of this study, the results have implications for the link between transitivity and morphological case in general. Following Hopper & Thompson (1980), we regard transitivity as a semantic-pragmatic property of clauses, rather than referring to the syntactic structure of verbs or the number of their arguments. Following Koenig et al. (2003), Rissman et al. (2015), we regard predicate argument structure as a semantic, scalar property of verbs, as syntactically expressed through governed case and semantically through event participanthood.

Section 2 provides background, outlining a number of approaches to oblique complements and discussing both indirect and direct semantic effects. Section 3 describes the method for compiling the datasets based on an Estonian corpus, while Section 4 introduces the design for investigating stativity, semantic roles and productivity. Section 5 presents the results, which are discussed in Section 6.

## 2. BACKGROUND

Oblique complements have become a topic of interest because of their unclear argument status. Arguments are often defined in opposition to adjunction, the two together forming the class of verbal dependants. Arguments are thought to fill in a verb's lexical frame, constituting its core participants which may or may not be syntactically realised in a clause ('I ate A PIE'). Adjuncts, on the other hand, are thought to express more peripheral information about the event, containing information which does not necessarily follow from the semantics of the verb (Koenig et al. 2003, Ackema 2015), for example, 'I painted this room FOR MUM'.

In addition, arguments are commonly marked with semantically bleached cases, i.e. structural cases. On the one hand, such cases are expected to be found with a wide range of semantically unspecific material, and on the other hand, they ought to lack specific semantics themselves, marking solely the syntactic relation between the complement and the verb. These two requirements have contributed to a view of oblique complements as non-arguments or semi-arguments, because they either constitute a semantically specific class (e.g. Experiencers), or the cases marking them appear to 'bring along' specific semantics. Broadly, these two distinctions can be viewed as indirect and direct semantics (Nichols 1983).

### 2.1. *Indirect semantics*

Non-canonical case marking is cross-linguistically associated with a variety of phenomena not necessarily related to the particular semantics of the chosen case (Nichols 1984, Onishi 2001, de Hoop & Malchukov 2007, Malchukov & De Swart 2008, Kittilä 2008, Lestrade 2010). A case can become linked to a particular semantic role, such as Experiencers ('THE CHILD is sad'.), or other less agentive participants ('YOU know this'.). A case can also become linked to particular verb types, such as modal verbs ('need'), verbs expressing occurrences ('drift') or verbs

of possession, existence and lacking ('have', 'lack', Onishi 2001: 25). In an attempt to systematise these patterns, Tsunoda (1981) proposed a verb type hierarchy. Malchukov (2005) postulated an elaborated version of Tsunoda's hierarchy, where some verb classes are more likely to take obliques due to the decreased subjecthood of their agents (e.g. perception and emotion verbs), while others select obliques due to the decreased patienthood of their objects (e.g. pursuit verbs; Malchukov & De Swart 2008: 343). These hierarchies are based on semantic features indicating transitivity (Hopper & Thompson 1980, Tsunoda 1981) and the Proto-Agent and Proto-Patient properties in Dowty (1991).

### 2.1.1. *Stativity*

While the transitivity parameters of Hopper & Thompson (1980) have been helpful in investigating the reasons behind non-canonical case marking in general, they include a great deal of covariation (Malchukov & De Swart 2008: 340). The covariation points to an underlying commonality between parameters. One feature underlying several of the transitivity parameters is stativity. As a lexical property of verbs, stativity broadly refers to two different properties, non-dynamicity and atelicity of states of affairs, meaning that participants do not move, nor do the events themselves have inherent temporal endpoints (Vendler 1957, Lakoff 1966, Onishi 2001). However, in its classic Aktionsart sense, this category is even narrower: stative verbs describe non-volitional states-of-affairs, which is why they lack continuous form (Vendler 1957: 146), and are ascribed to the grey area between properties and actions (Vendler 1957: 152). Hence, such volitional, non-dynamic, atelic verbs as 'lie' and 'sit' are often excluded from the group of stative verbs.

Indeed, stativity has received somewhat more attention as a transitivity parameter and appears to play a fundamental role, possibly explaining the covariation of a number of other parameters. It is included in Hopper & Thompson (1980) as the parameter 'kinesis' with the binary values of actions versus states, but it is also entailed in other parameters, for example, punctuality, volitionality, agency, affect-edness of object, and it is related to the individuation of object and aspect. Tsunoda (1981: 396) sees stativity as underlying his entire transitivity-based verb type hierarchy, constituting the variable measured by other features. Onishi (2001) mentions stativity as one of the three main factors affecting verbs' case selection, along with lack of control (volitionality) and modality (irrealis vs realis). Hence, we are justified in hypothesising that stativity is more fundamental to this matter than any other lexical semantic factor. This makes it a useful starting point for an empirical study of the semantics of verb complements in spatial cases.

Yet there remains some confusion around the concept of stativity, originating in uncertainty about whether states should be viewed in the narrower sense of Vendler (1957) as properties and relationships, or in a broader sense, according to which states also include volitional, non-dynamic, atelic verbs, such as 'sit'. In an approach developed by Maienborn (2003), the two types are joined together into

a single framework. In this approach, statives form a multilevel system (Maienborn 2003, 2008, Rothmayr 2009, Maienborn 2019), divided into K-states (or Kimian states, following Kim (1969)) and D-states (or Davidsonian states, following Davidson (1967)). K-states ('own', 'love') constitute abstract relationships and properties rather than actual events. They have no location in space but can be located in time (Maienborn 2008: 113). K-states, therefore, broadly align with stative verbs in Vendler (1957) and Lakoff (1966).

D-states, on the other hand, differ from K-states in that they meet the criteria for events in the sense of Davidson (1967). These represent physically stative, atelic events ('sit', 'lie', 'wait'), which, unlike K-states, exist in both space and time. K-states can be determined by means of several 'tests': for instance, they are expected to resist regular deverbal nominalisation morphology (Maienborn 2003, Rothmayr 2009), for example, \*'the resembling', and they do not occur as complements of perception verbs, as they cannot be perceived, for example, \*'I saw the tomatoes weigh 1 pound' (Maienborn 2019: 41). Verbs which are neither K-states nor D-states are lexically dynamic and/or telic, meaning they have natural time boundaries (Fábregas & Marín 2012). All in all, this framework is more inclusive and internally gradient, linking the class of true property-like states (K-states) to the rest of the verb space through the grey area of non-dynamic, atelic events (D-states).

In addition to transitivity parameters, many approaches highlight aspect as one of the main mediators between semantics and syntax (Tenny 1992, Verkuyl 1999, Kratzer 2004). For instance, Tenny's 1992 approach defines internal arguments through aspect, proposing that they 'measure out the event' (through change), sometimes also delimiting events (encoding their temporal end points, e.g. 'destroy') but not always ('push'). However, Tenny's study does not address English complements marked with adpositions, which likely mostly overlap with the non-canonically marked complement verbs in the present study (Nichols 1984). If, however, such verbs were shown to systematically differ from canonical complement verbs in terms of lexical aspect, this could be regarded as support for aspect-based approaches to argument realisation. Their lack of natural time boundaries could then be viewed as a likely factor contributing to their distinct morphological marking.

Returning to indirect semantic effects, they do not only refer to the indirect, yet specific meanings which become linked to oblique case usage in complement marking but also to the potential motivations for a language having a complex system of complement marking in the first place. Revealing patterns unique to oblique complement verbs, in comparison with canonical complement verbs, may provide a key to understanding the function of oblique case-marking for verbal complements.

Stativity constitutes a suitable initial axis for mapping indirect semantic effects in this comparison. In general, but especially within Maienborn's approach (Maienborn 2003, 2008), stativity resists being conceived of as merely one of the transitivity factors, as it is inherently linked to three of the four lexical semantic parameters of verbs in Hopper & Thompson (1980): kinesis (dynamicity), punctuality and volitionality, leaving aside only the number of event participants.

Moreover, the graded view of stativity allows us not only to use K-states to investigate the role of states in the sense of Vendler (1957), i.e. that of properties and relationships, but also to use D-states and the decomposed notions of dynamicity and aspect to investigate the role of transitivity. If aspect plays a central role in syntactic realisation, as expected by the aspectual hypothesis of syntactic realisation (Tenny 1992), we would expect it to surface as an important distinction between oblique and canonical complement verbs.

## 2.2. *Direct semantics*

Differences between oblique and canonical complement verbs also arise through patterns of direct semantics. Non-canonical marking can directly index the semantic role of an argument and/or verbal semantics (Nichols 1983: 186), with a transparent, direct semantic link. However, such effects vary to some degree, depending on the complement. For instance, many Estonian verbs alternate between concrete and abstract complements, the first of which resemble spatial adverbials, while the latter constitute a more abstract relation:

- (2) (a) kaamera läheneb näole  
 camera approach:3SG face.ALL  
 ‘The camera approaches the face’.
- (b) ma lähenen teemale teaduslikult  
 1SG approach:1SG topic:ALL scientifically  
 ‘I am approaching the topic scientifically’.

While the lative Goal semantics of allative case is present in both examples (Examples 2a–2b), the allative complement in Example (2b) is not a physical location while the allative element in Example (2a) is. Although the metaphorical presence of allative meaning in Example (2b) might lead some analysts to disallow TEEMALE as an argument, others would claim that spatial meaning is less clear with abstract complements.

As pointed out in Nichols (1983: 182), direct semantic effects can also arise diachronically. Several Russian verbs of negative emotion, for example, *IZBEGAT* ‘avoid’, allow genitive case to be used in addition to the canonical accusative, most likely because genitive is historically linked to the Indo-European ablative, denoting separative movement away from something.

Direct semantic effects also provide a lens for comparing oblique cases. Following the observation that the direct semantic effects of a case are more clearly present with some verbs than others, and that a spatial case can lack them altogether, as in Example (1), one may compare cases in terms of how frequently they are used with verbs when such effects are present. This is an understudied property of obliques, especially considering the insight it may provide into the internal organisation of the umbrella category termed ‘oblique complements’.



### 2.2.1. *Grammaticalisation cline*

Another property of individual cases which does not often feature in studies of non-canonical case use is the degree of grammaticalisation. Namely, it has been observed that semantic cases are linked to both prepositions and syntactic cases in a diachronic chain. This is represented by the case cline shown in Example (3), which maps a pathway from nouns to syntactic case markers (Bybee 1985, Hopper & Traugott 2003, Lestrade 2010).

- (3) Noun → Spatial adposition → Spatial case → Syntactic case  
(Lestrade 2010)

Each category on the cline tends to develop from the one on its left. For instance, the Estonian comitative suffix *-GA* originates from the comitative postposition *KAAS* (Rätsep 1979, Aigro 2020) and the Estonian syntactic object case, partitive, can be traced back to separative, a spatial case (Rätsep 1979).

The oblique complement-marking function is expected to occur only with some semantic cases, rather than all, because some can be expected to have higher degrees of grammaticalisation than others. Dative has been noted to function as an oblique subject case in Icelandic (Zaenen & Maling 1984) and German (Haspelmath 2001). The same is true for genitive in Bengali (Onishi 2001). Instrumental has been noted to be an oblique object case in Russian, allative in Chechen-Ingush, dative in Nanai (Nichols 1983). Out of the 11 semantic cases in Finnish, Nichols regards seven as sometimes marking argument roles (Nichols 1983: 175).

### 2.2.2. *Grammaticalisation and productivity*

In order to assess the assumed variance between spatial cases, one ought to assess and compare their individual degrees of grammaticalisation. For the purposes of this study, productivity is a useful way to quantify grammaticalisation.

Productivity in complement structures can be investigated from two different perspectives: the morphological productivity of case affixes and the syntactic productivity of structures where case occurs in a selected position. A number of quantitative methods have been proposed for measuring morphological productivity (for a comprehensive review, see Zeldes 2012). The predominant methods include realised, expanding and potential productivity (Baayen 1993, 2009, Zeldes 2012).

Realised productivity refers to the ‘area’ the morpheme currently covers, measured by type count, that is the number of lemmas a morpheme is attached to in a particular corpus or dataset. Type count is a common measure of productivity (Zeldes 2012, Barðdal 2008), reflecting the number of lemmas or types to which the form applies, and it increases upon moving right on Example (3). Syntactic cases are compatible with a larger number of distinct lexeme types than semantic cases, meaning they have a higher type count. The grammaticalisation of a semantic case



into a syntactic case inherently entails an increase in its productivity and, therefore, type count.

This measure only describes the past production of a morpheme, but grammaticalisation involves change over time, and so other measures are needed to predict its future use. For this purpose, Baayen (1993) proposed using HAPAX LEGOMENA. Hapaxes are lexemes detected with the relevant morpheme only once in a corpus. For instance, the lemma *UIGUUR* ‘a person of Uyghur descent’ only has one elative token in the Balanced Corpus of Estonian (*UIGUURIST*), meaning that it constitutes a hapax from the perspective of elative. Hapaxes are used to calculate expanding and potential productivity. Expanding productivity requires dividing the morpheme’s hapaxes with the number of all hapaxes in a corpus. Measuring potential productivity requires dividing the morpheme’s hapaxes with the number of all tokens of that morpheme.

When using a dataset of sentences instead of an entire corpus, potential productivity is more easily used. This means dividing the hapaxes formed with a particular case (e.g. the number of lemmas with only one elative token) with the number of tokens of that case (all elative tokens in the corpus). Simply put, the larger the proportion of hapaxes in a morpheme’s total uses, the more potential it may be said to have in terms of being extended to new lexemes. In a dataset where each case is equally represented, the divisor is constant, meaning that potential probability will be indicated via simple hapax count.

With regard to syntactic productivity, it works in an analogous way to morphological productivity. Phrases are seen as analogous to word forms, using recurrent structural parts and variable lexical elements (Zeldes 2012: 23). Barðdal (2006) views a particular kind of selected case construction, for example, the ditransitive construction, as a recurrent part, and the verb types occurring in it as the variable elements. Non-constructional approaches sometimes call such a construction a case frame (Malchukov & De Swart 2008). A case frame can be said to have higher syntactic productivity if more verbs occur in it, making it more likely that speakers will use the frame with novel verb types (Barðdal 2008: 29). More grammaticalised cases are expected to form syntactically more productive case frames.

In summary, higher syntactic and morphological productivity in cases indicates a higher degree of grammaticalisation. If more grammaticalised cases can also be shown to mark direct semantic effects less frequently than less grammaticalised cases, this would go against the mainstream view which regards direct semantics as a core property of oblique complement verbs, distinguishing them from canonical complement verbs. Instead, direct semantic effects would look more like a system of case-based tendencies, where some groups of oblique complement verbs are considerably more similar to canonical complement verbs than others. A quantitative approach is crucial for an exploration of this system, as this will give us a more complete picture of the usage of verbs and the range of functions (structural and semantic) in which cases are used, than a study built on qualitative analysis of individual examples.

### 2.3. *Estonian cases and spatial complements*

Estonian has a rich morphological system, with 14 suffix-based morphological cases, three of which are regarded as syntactic ('abstract') and 11 as semantic ('concrete', Viht & Habicht 2019). The semantic cases include six spatial cases: three external (allative 'onto', adessive 'on', ablative 'from on') and three internal locative cases (illative 'into', inessive 'in', elative 'from in'). In addition, terminative ('up until') is also regarded as having some degree of spatial functionality (Erelt et al. 1995).

#### 2.3.1. *Canonical arguments*

The vast majority of subjects are in nominative, which is morphologically unmarked. However, differential subject marking occurs between nominative and partitive in existential clauses, where partitive seems to indicate reduced transitivity and less agentive subjects (Lindström 2017). In addition, some Experiencers occur in allative or adessive, while some verbs (e.g. OLEMA VAJA 'need') show differential marking between the two, the choice of case depending on a number of factors (Lindström & Vihman 2017). A comprehensive description of subject-like elements in Estonian can be found in Metslang (2013).

For objects, the situation is more complicated. Differential object marking involves partitive, genitive and nominative (Rajandi & Metslang 1979, Erelt et al. 1993), as in Examples (4–6):

- (4) Maarja sõi leiba  
 Maarja eat.3SG.PST bread.PAR  
 'Maarja ate bread'./'Maarja was eating bread'.
- (5) Maarja sõi leiva  
 Maarja eat.3SG.PST bread.GEN  
 'Maarja ate the whole bread'./'Maarja ate the bread up'.
- (6) Söö leib ära!  
 eat.IMP bread.NOM up  
 'Eat this bread!'

This paradigm applies to a much wider range of contexts than differential subject marking, and, hence, it is regarded as the core definition of syntactic relations in Estonian (Erelt 1989: 15). At the same time, the choice of case depends on various semantic and pragmatic factors (e.g. aspect, quantitative specificity, number), analysed through transitivity parameters in Hiietam (2003). The extent of semantic and pragmatic effects has led Miljan (2009) to argue that partitive and genitive, too, constitute semantic rather than syntactic cases, although encoding underspecified rather than truth-conditional semantics.

#### 2.3.2. *Obliques*

Estonian spatial case-marking has received a fair amount of attention in the context of event participant marking. Spatial cases are rather multifunctional, marking

spatial, temporal, possessive, causative elements, as well as event participants. Non-canonical subjecthood, for example, Experiencers, is associated with adessive ('on') and allative ('onto'). Overall, subject-like elements (as measured via the Proto-Agent properties in Dowty (1991), e.g. animacy) are common in modal (Kehayov & Torn-Leesik 2009, Lindström & Vihman 2017), possessive (Erelt & Metslang 2006) and passive constructions (Lindström & Trigel 2010). All in all, allative and adessive appear to correspond to the Indo-European dative in many respects (Matsumura 1994), both in terms of non-agentive subjects as well as Recipient (indirect object) and Beneficiary roles, often marked by these cases. In a comparison of adessive and allative in the 'need' construction, adessive has been suggested to be more grammaticalised and abstract than allative (Lindström & Vihman 2017).

Other spatial cases have been less discussed, perhaps with the exception of elative case. As a highly multifunctional marker, elative is used across an unusually broad semantic field (Saluveer 1958, Nurka 2014), part of which is far from its original spatial meaning. (Saluveer 1958) identifies 17 semantic functions for elative, including time, agency and relationship. There is, however, no systematic review of elative or other spatial cases marking event participants, nor has abstractness in this role been quantified or linked to their grammaticalisation.

There is, however, a nearly century-old debate over the argument status of various elements marked in spatial cases. The predominant view has always followed Kettunen (1924) in classifying obliques as a subclass of adverbs (Valgma & Remmel 1968, Mihkla et al. 1974, Erelt 1989, 2003, 2004, Veismann et al. 2017). As one of the few researchers representing the dissenting opinion, Kure (1959) claimed that the inessive-marked complement of the verb *KAHTLEMA* 'doubt' constitutes a type of object. He noted that, if it is accepted that nominative case can mark both subjects and objects, it should also be accepted that inessive can mark both adverbs and objects. Similarly, Klaas (1988) gives examples of various semantic cases marking object-like elements (e.g. the illative complement of *HUVITUMA* 'be interested in'). However, neither regards oblique elements as entirely analogous to canonical objects, one of the main obstacles being that they do not participate in the Estonian differential object marking system (Klaas 1988). However, governed spatial case is included in the valence system of Rätsep (1978) in a manner that is analogous to canonical DOM cases.

#### 2.4. *This study*

This paper focuses on Estonian verbs selecting complements marked in spatial case. These verbs are sometimes called 'verb splits', that is verbs with a fixed oblique case, rather than depicting case alternations between several cases ('transitivity alternations', Malchukov 2005). As a morphologically rich language, Estonian is especially well-suited for investigating structural case functions in languages with a

nominative-accusative alignment. It has a wide range of cases which have been noted to have both semantic and structural functions, of which this study tackles six.

As the paper is focused on verb splits rather than case alternations, it investigates lexical semantics rather than clause variables, such as tense, aspect or mood. We consider such verbs fruitful for investigating the morphosyntactic marking of event participants, following the notion in Nichols (1984) that it could be more revealing to focus on the peripheral areas of verbal dependency rather than merely its prototypical instances. In this case, ‘peripheral’ refers to non-canonical case marking, that is semantic cases used in structural functions, while ‘prototypical’ refers to arguments in structural cases and adjuncts in semantic cases.

Verbs in our dataset were extracted by means of verb-complement co-occurrence frequency. While the dataset compiled by this means does not pretend to be comprehensive, it is a useful means of extracting data from a corpus, free from semantic bias. This allows us to investigate the lexical semantic motivations underlying non-canonical marking in a dataset compiled by non-semantic measures.

We aim to answer the following research questions:

1. Which indirect semantic effects underlie the distinction between oblique and canonical complement verbs?
2. Do direct semantic effects differ by case, and how are these differences linked to the degree of grammaticalisation of a case?

The first question is addressed in [Section 5.1](#), where we take a detailed look at the role of stativity. We hypothesise that there is a correspondence between stativity and case that affects all levels, including K-states (‘depend on’, ‘weigh’), D-states (‘sit’, ‘wait’) and non-states. The second question is addressed in [Sections 5.2–5.3](#), first by comparing spatial cases in terms of the semantic roles they mark, and second, linking their role patterns to their individual degrees of grammaticalisation. Based on the literature, we hypothesise that: 1) elative and allative entail direct semantic effects less frequently than other spatial cases when marking complements, and that 2) they are more grammaticalised than the other cases. This study is the first broad, data-based characterisation of this phenomenon in Estonian, simultaneously investigating two different semantic dimensions of obliques.

### 3. DATA

Two datasets are analysed in this study, broadly focusing on four properties: stativity, semantic roles, morphological and syntactic productivity. Stativity is analysed by comparing datasets comprising oblique complement verbs and canonically transitive verbs. The other three properties are investigated in only the oblique complement dataset. Stativity and semantic role data is freely available (Aigro 2022).

The oblique complement verb dataset and the canonical transitive verb dataset were both obtained from the Balanced Corpus of Estonian, a morphologically parsed 15-million word corpus of written Estonian, consisting of equal samples of texts from journalism (daily and weekly newspapers, 1995–2007), science (dissertations and articles from various disciplines, 1995–2006) and fiction (prose and poetry, 1987–2011).

### 3.1. *Oblique complement data*

Our first dataset represents oblique complement verbs extracted using their complement co-occurrence frequencies. Using Python 3.6., we mapped how frequently each Estonian verb co-occurs with each of the six spatial case suffixes in a single subclause. This resulted in six percentage values for each verb. We then extracted all clauses containing verbs which: (a) co-occur with a single spatial case in more than 50% of their instances and (b) are used with a relative frequency exceeding 1 per million words (pmw).

There is some debate about the role of co-occurrence frequencies in defining argumenthood, including questions regarding whether argumenthood is defined by obligatory syntactic expression (Barbu & Toivonen 2016), whether high co-occurrence frequencies are merely one of the factors contributing to prototypical argumenthood (Arka 2014, Forker 2014) or whether they are not central to (semantic) argumenthood, which is based on event participant status instead (Koenig et al. 2003, Rissman et al. 2015). However, this paper does not address the argument status of the verbal dependants involved. Hence, the only relevant questions are: will co-occurrence frequencies render oblique complement verbs, and does this method introduce some type of semantic bias into the datasets?

The answer to the first question is affirmative. It is a useful data extraction method, not only because it renders the necessary data, but also because it has the advantage of not being inherently semantic. This is crucial if we wish to investigate whether the two datasets differ in semantics. The second question is more complicated. While Estonian has argument ellipsis on all levels, subjects are especially likely to be omitted, both in spoken and written Estonian and with both canonical and non-canonical case marking (Lindström & Vihman 2017). Hence, verbs with non-canonical subjects (e.g. Experiencers) are potentially less likely to be represented than other verbs. This would most likely affect the proportion of verbs governing adessive and allative, that is the two cases linked to Experiencer arguments.

The dataset of all instances of verbs with sufficient frequency and complement co-occurrence rate was manually cleaned, removing infinitive verb forms, adjectival past participles, wrongly annotated non-verbs and wrongly annotated spatial case. All clauses where the spatial element was not automatically detected ( $n = 10,693$ ) were manually checked. We then used the collection of sentences including complements ( $n = 36,216$ ) to compile a smaller sample ( $n = 5,234$ ) with just 50 random examples per verb — or the total number of clauses for verbs with less than 50 instances (mean instances per verb = 42.2,  $n = 5,234$ ). We used the sample

to confirm the reliability of our complement co-occurrence rate data. Only 1.6% of clauses automatically annotated as including complements did not actually have them, meaning the co-occurring spatial case phrase was not a complement. This was deemed an acceptable error rate for the purposes of this study. The data of verbs for which the error rate exceeded 5% were manually cleaned in full.

Finally, we observed that some dataset verbs appeared to co-occur frequently with physical locations (e.g. *PATRULLIMA* ‘patrol’). Although such elements constitute an interesting subtopic in argumenthood, they do not fit in the present study. We excluded 18 such verbs from the oblique complement dataset by means of two variables: the semantic role and the animacy of the spatial case-marked element. The former was coded as a verb-based variable (see [Section 4.2.1](#)), while the latter was coded in the clauses in the small sample ( $n = 5,234$ ), using four categories: human animate, other animate, inanimate concrete and inanimate abstract. A verb was excluded from the dataset if more than half of its complements simultaneously: 1) had the semantic role of Goal, Source or Location, and 2) were coded as inanimate concrete entities. For instance, this included *LAHUSTUMA* ‘become dissolved in’ and *NÕJATUMA* ‘lean on’.

Although useful for the purposes of this study, such artificial cut-off points as the 50% requirement unfortunately cover up the inherent gradience of the research object. For instance, *ERALDUMA* (‘separate from’) remains in the oblique complement dataset, even though it co-occurs with physical locations, for example, gas emanating from factories, in 45% of its tokens. The rest are inanimate abstract Sources, Example (7), where the verb’s meaning is close to ‘differ from’:

- (7) *kultuur eraldub kunstist*  
 culture separate:3SG art.EL  
 ‘Culture separates from art’./‘Culture becomes distinguished from art’.

In most of the dataset, physical locations are infrequent. Only 10% ( $n = 12$ ) of the dataset verbs co-occur with a physical location somewhat frequently (30%–50% of their tokens).

The final dataset includes 104 oblique complement verbs (see [Table 1](#)). We intentionally set no restrictions on the types of verbs, allowing the dataset to contain two-place verbs with Experiencers, Example (8a), as well as two- and three-place verbs with object-like Themes, Examples (8b–8c).

- (8) (a) *tal ei õnnestunud saada soovitud haridust*  
 3SG:AD NEG succeed:CNG.PST get:INF wanted education:PART  
 ‘S/he did not succeed in obtaining the education they wanted’.
- (b) *ma ei hooli eriti sünnipäevadest*  
 1SG NEG care:CNG much birthday.PL:EL  
 ‘I don’t care much about birthdays’.

Case	Number of verbs	Number of observations (% of dataset)	Complement rate
Allative ('onto')	39	17168 (39.3%)	78.1%
Aessive ('on')	5	4044 (9.2%)	67.5%
Illative ('into')	12	2133 (4.9%)	77.9%
Inessive ('in')	10	4309 (9.9%)	74.9%
Elative ('out of')	38	16040 (36.7%)	83.8%
Total	104	43694	78.9% (mean)

*Table 1*  
Oblique complement dataset.

- (c) psühhiaater süüdistas stressi kasvus elutempo  
 psühiatrist accuse:PST stress.GEN increase:IN life pace.GEN  
 kiirenemist  
 acceleration.PART  
 'The psychiatrist blamed the increase in stress on the accelerating pace of life'.

In addition to the three spatial cases in the previous examples, oblique complements can also occur in illative and allative, Examples (9a–9b):

- (9) (a) ma ei usu imedesse  
 1SG NEG believe.CNG miracle:PL:ILL  
 'I don't believe in miracles'.  
 (b) kolmapäevale järgneb neljapäev  
 Wednesday:ALL follow:PRS Thursday  
 'Thursday follows Wednesday'.

No verb co-occurred with ablative ('from on top of') elements frequently enough to be included in the dataset. [Table 1](#) describes the oblique complement dataset:

### 3.2. *Canonically transitive data*

The second dataset is made up of verbs that occur frequently and take complements in syntactic cases, that is partitive, genitive or nominative (verbs with a relative frequency of over 100 pmw (n = 260)). Transitive verbs could not be automatically selected based on overt object frequencies, because due to the multiple functions of partitive, genitive and nominative, the occurrence of these cases does not imply the occurrence of a direct object. All object occurrences had to be manually coded.

Verbs were extracted with their frequency of co-occurrence with partitive and genitive elements. Although the Estonian DOM paradigm also allows for nominative direct objects, it is the least frequent object case and the most common subject case. Hence, it was excluded from the automatic extraction. We then compiled a



sample of clauses including verbs with the highest cumulative partitive and genitive co-occurrence rates (50 instances per verb, 222 verbs, 11,100 observations). The presence/absence and case of direct objects was manually coded in the sample, followed by a further exclusion step. The final dataset only includes verbs co-occurring with objects in more than half of their instances (128 verbs, 6,400 observations).

#### 4. METHOD

##### 4.1. *Method for addressing indirect semantic effects*

In order to answer our first question about the semantic motivations underlying the non-canonical marking of verb complements, we compared the degree of stativity in oblique complement verbs to that of canonical transitive verbs. For this purpose, verbs in both datasets were tagged as K-states, D-states or non-states.

We extracted K-states by means of an acceptability judgment task and three different semantic variables. We planned to also include nominalisation frequencies as evidence for encoding K-states. However, the suffix *-MINE* is the most productive deverbal nominalisation suffix in Estonian (Kasik 2015, Pilvik 2019). Our corpus search revealed that oblique complement verbs actually showed higher nominalisation frequencies than canonical complement verbs, when we controlled for the frequency of verb stems. Due to the varying productivity of nominalisation morphology across languages, this might not constitute a helpful cross-linguistic test for identifying K-states.

Instead, we coded K-states based on: 1) an acceptability judgment task, 2) semantic verb type, 3) dynamicity and 4) lexical aspect. The judgment task constituted the broadest filter.

##### 4.1.1. *Acceptability judgment task*

Based on the observation that K-states only exist in time and not in space, it has been suggested that they cannot be perceived (Maienborn 2008, Rothmayr 2009), and that therefore, verbs expressing them do not occur as perception verb complements. We tested the acceptability of Estonian oblique and canonical complement verbs acting as complements of the perception verb *NÄGEMA* ‘see’. Participants used a 7-point Likert scale to assess sentences which all had same structure as Example (10):

- (10) Ma nägin            Indrekut        Kallele        virutamas  
 1SG see:1SG.PST Indrek.PART Kalle:ALL hit:PROG  
 ‘I saw Indrek hitting Kalle’.

The visual modality was chosen as the primary sense modality expressed in Estonian, vision verbs being the most frequent perception verbs (Proos 2021). In order to keep other parts of the clause as acceptable as possible, all elements besides the main subject (1SG), the finite verb and the infinitive were chosen based on the

highest collocation log-likelihood.<sup>2</sup> The experiment included 232 experimental items (one item per verb, 104 oblique complement verbs and 128 canonically-marked complement verbs). Experimental items were divided into two lists, with 49 fillers added to list 1 and 48 fillers added to list 2. Participants included adult L1 Estonian-speaking participants, who completed the survey in a web application. List 1 was completed by 26 participants and list 2 by 17 participants. Verbs appearing in sentences with average ratings of less than 4 points on the 7-point scale were considered possible K-states.

#### 4.1.2. *Verb type*

Verb type was coded with the levels: 1) Experiencer/Possessor verbs (HOOLIMA ‘care about’, KUULUMA ‘belong to’), 2) measure verbs (KOOSNEMA ‘consist of’, KUBISEMA ‘be abundant in’), 3) properties (RAJANEMA ‘be founded on’, VÖLGNEMA ‘owe’) and 4) other. The former two are discussed in Rothmayr (2009) as verbs expressing features of K-states, while the third level is part of Maienborn’s core definition of K-states: they constitute properties of entities (Maienborn 2008: 113).

#### 4.1.3. *Dynamicity and aspect*

Dynamicity and lexical aspect were both coded as binary variables. A verb was deemed dynamic if it had at least one sense in which the event requires one or all participants to physically move, even if only slightly (LÄHENEMA ‘approach’ is dynamic, while HOOLIMA ‘care about’ is not). A verb was deemed telic if it entailed an inherent temporal end point in at least one of its senses (KUULETUMA ‘obey’ is atelic, while LOOBUMA ‘give up’ is telic).

#### 4.1.4. *Coding states*

A verb was coded as a K-state if it met four conditions: 1) not accepted as the complement of a perception verb, 2) tagged as an Experiencer, Possessor, measure or property verb, 3) non-dynamic and 4) atelic.

The rest of the verbs were then examined in terms of dynamicity and telicity. Following the definition that D-states constitute spatiotemporal eventualities which entail neither motion nor temporal endpoints (Rothmayr 2009), a verb was coded as a D-state if it was: 1) not a K-state, 2) non-dynamic and 3) atelic (e.g. TOETUMA ‘lean against’/‘be supported by’). The rest of the verbs were coded as non-states.

Encoding semantic variables introduces fuzziness into the design, which raised three issues. First, a binary encoding of dynamicity and aspect naturally ignores the gradience of these variables. Second, some verbs entail more senses than others, which render their lexical forms ambiguous in various dimensions. Ambiguity can arise from subjects (Rothmayr 2009), for example, with an inanimate subject JÄRGNEMA ‘follow’

[2] Data originate from the collocation database of the Estonian National corpus: <https://korpused.keeleressursid.ee/clc/?mod=search>.

is stative and non-dynamic, while it is non-stative and dynamic with an animate subject (ESMASPÄEVALE JÄRGNEB TEISIPÄEV ‘Tuesday follows Monday’ versus POISS JÄRGNEB KOERALE. ‘The boy follows the dog’). Ambiguity can also arise from the complement, for example, LÄHENEMA ‘approach’ is dynamic with a concrete complement and non-dynamic with an abstract complement (KALLE LÄHENEB RANNALE/PROBLEEMILE ENESEKINDLALT. ‘Kalle approaches the beach/the problem confidently’). Third, a number of verbs are highly abstract, their stativity, dynamicity and aspect relying almost entirely on their complements, including verbs which can take deverbal nominalisations (JÄTKAMA LUGEMIST ‘continue reading’) or infinitives (SUNDIMA LUGEMA ‘force (so) to read’) as complements. All three issues were resolved by opting for a conservative approach. A verb was coded as non-stative and dynamic, even if it entailed very little movement (interaction verbs), if it had at least one dynamic sense, lexically or with a nominalised/infinitival complement.

#### 4.2. Method for addressing direct semantic effects

Direct semantic effects were investigated through semantic roles. In order to examine their link to the degree of grammaticalisation of each spatial case, these degrees were mapped by way of the morphological and syntactic productivity of spatial cases.

##### 4.2.1. Semantic roles

We used the roles of Source, Location and Goal to mark the presence of spatial semantics in complements. We used Experiencers, Recipients and Themes to mark complements without spatial meaning. No other roles were used. Only complements in spatial cases were assigned roles. Roles were independently coded by both authors of this paper, as well as a third researcher. Inter-coder agreement was 97.1%.

Encoding roles marked by complements in non-canonical cases is complicated by definition, because as discussed in Section 2.1.1, these complements are expected to show decreased subjecthood and objecthood compared to prototypical arguments of canonical transitive verbs. Semantic roles were originally designed to refer to entities (Actor, Patient, Theme, Instrument) and the spatiotemporal properties of these entities (Goal, Source, Location, Path) (Klçaslan & Tuna 2015). These labels are not easily transferrable to verbs describing abstract relationships between often inanimate entities (e.g. ‘The budget depends on my income’). Hence, we took a prototype perspective to semantic roles (Hopper & Thompson 1980, Dowty 1991, Primus 1999, Rissman & Majid 2019), where roles only have very broad definitions. Concrete or abstract entities involved in relationships, as well as properties of various entities, were viewed as Themes, for example, the dependee of ‘depend’.

##### 4.2.2. Productivity

This study measures both morphological and syntactic productivity. We assessed the morphological productivity of spatial cases by means of two statistics — realised

productivity (type counts), that is the number of lemmas occurring with an affix, and potential productivity, that is the proportion of a morpheme's HAPAX LEGOMENA in all its tokens (Baayen 1993). These two measures were used on a dataset of lexemes obtained from the dataset of oblique complement verbs. The two variables were measured in five random samples of 627 case-marked lexemes, representing five spatial cases marking complements (when extracting relevant case-marked lexemes from 500 clauses, 627 was the smallest number of case-marked lexemes occurring per case).

The number of verbs governing a particular case can be regarded as a measure of the syntactic productivity of that case frame. This variable is mapped solely in the oblique complement dataset, which is not entirely bias free (as discussed in Section 3.1). However, it is sufficient for the purposes of this study, as it is only one component in the more complex productivity variable.

## 5. RESULTS

### 5.1. *Stativity*

Stativity was analysed as a categorical variable with three levels: K-states, D-states and non-states.

#### 5.1.1. *K-states*

The perception verb test described in Section 4.1.1 examined the acceptability of verbs in both groups as complements of the perception verb NÄGEMA ('see').

As shown in Figure 1, oblique complement verbs are significantly less acceptable when they occur as perception verb complements, compared to canonical complement

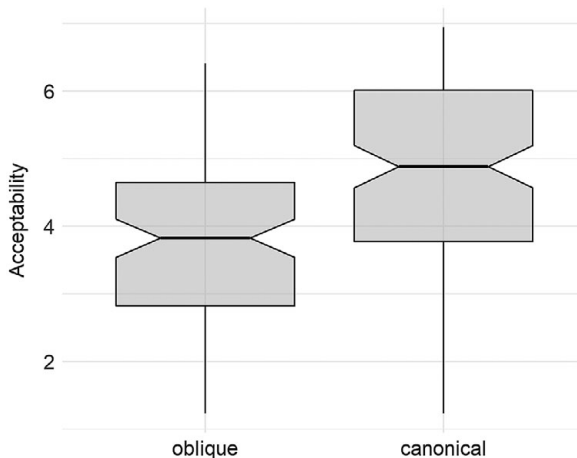


Figure 1

Acceptability judgments of oblique and canonical complement verbs when occurring as perception verb complements.

verbs (oblique: mean = 3.9, sd = 1.3, on a 7-point scale; canonical: mean = 4.8, sd = 1.3,  $W = 9,158.5$ ,  $p < .001$ ). The oblique complement dataset contains almost twice as many verbs with a mean rating below 4 than the transitive dataset (56% vs 30%).

Of the 58 oblique complement verbs with a mean acceptability judgment of less than 4, 37 verbs constituted Experiencer/Possessor, measure or property verbs, 31 of which were also simultaneously non-dynamic and atelic and were therefore coded as K-states. The same conditions rendered significantly fewer K-states in the canonical verb group ( $n = 12$ ,  $\chi^2 = 14.541$ ,  $p = .0001$ ).

### 5.1.2. D-states and non-states

D-states, that is non-dynamic atelic verbs, constitute 10% of the oblique complement dataset and 9% of the canonical complement dataset, which is not a statistically significant distinction. However, among non-states, non-dynamic verbs make up a larger proportion of oblique complement verbs ( $n = 17$ , 34%) than the canonical set ( $n = 10$ , 10%,  $\chi^2 = 11.637$ ,  $p = 0.0006$ ). The distinction is even greater when one considers that oblique complement verbs coded as dynamic are primarily used non-dynamically, as described in Section 3.1. For instance, in more than half of its corpus instances, LÄHENEMA ‘approach’ is used with abstract entities (e.g. ‘approach the truth’), giving it a non-dynamic interpretation.

The other feature contributing to stativity, that is lexical atelicity, is not as pronounced among the oblique dataset non-states as non-dynamicity. Most non-states in both datasets are telic (68% vs 53% in the canonical dataset), and there is no significant distinction between these two proportions.

Overall, oblique case frames do occur with verbs expressing K-states to a significantly greater extent than canonical case frames. Outside of that group, they

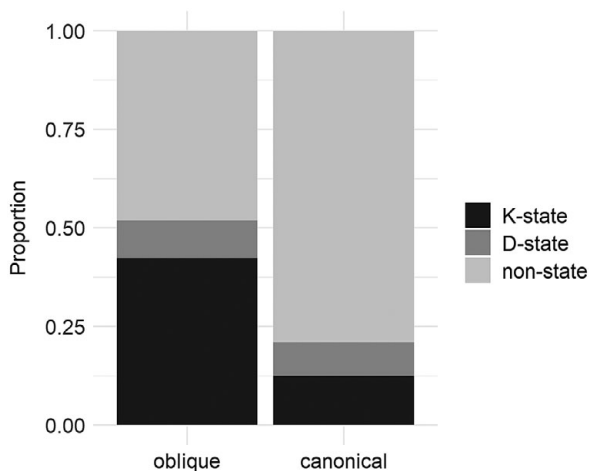


Figure 2

Stativity types in oblique and canonical complement verbs.

also differ from canonical complement verbs in terms of being more linked with non-dynamic verbs, while the two groups are somewhat similar in terms of aspect. The proportions of K-, D- and non-state verbs in the two datasets are summarised in Figure 2:

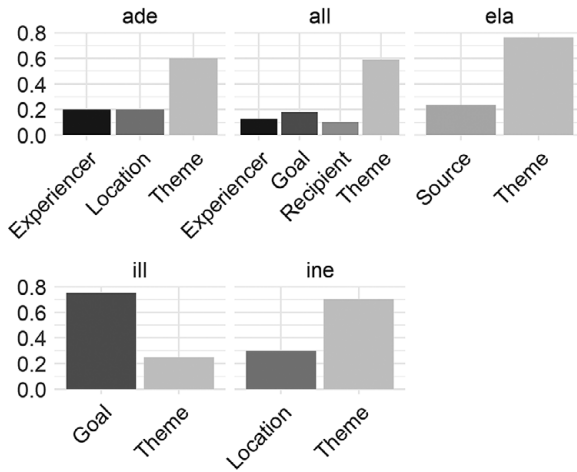


Figure 3

Semantic roles of oblique complements per case frame (% of verbs selecting these cases).

### 5.2. Semantic roles

In the context of direct semantic effects, most spatial cases entail spatial semantics in a minority of their instances, when marking verbal complements, as shown in Figure 3:

Figure 3 includes both roles entailing spatial semantics (Goals, Sources, Locations), as well as roles with no spatial semantics (Experiencers, Recipients, Themes). It appears that four spatial cases encode direct semantic effects rather infrequently, but this can be taken to imply a more generalised use only for allative and elative, which encode Goals and Sources in 18% and 24% of verbs, respectively. While adessive and inessive complements are infrequently semantically spatial, verbs governing these two cases cannot be viewed as semantically generalised, because most verbs in the datasets of these two cases, including all adessive Themes, constitute near-synonyms. Half of the inessive verbs denote a relationship of expression (KAJASTUMA ‘be expressed in’, PEEGELDUMA ‘be reflected in’), with another 40% denoting containment (SISALDUMA ‘be contained in’, FIGUREERIMA ‘figure in’). All adessive Theme verbs denote a relationship where something is based on something else. For illative verbs, spatial semantic roles are marked more often than non-spatial roles.

For elative and allative, however, these results point towards more generalised use than for the other three cases. These two cases mark spatial roles less frequently than others. Allative marks a much wider range of semantic roles than others, including the roles of Experiencer and Recipient, in addition to Theme. Elative has the highest proportion of Themes, where it occurs as a highly generalised and bleached complement marker. All in all, these two cases mark direct semantic effects least frequently. In addition, these results imply they might be more grammaticalised than other spatial cases. The question of grammaticalisation will next be investigated via productivity.

### 5.3. Productivity

We investigated two types of productivity, morphological and syntactic, among five spatial case affixes when they occur as complement markers. Morphological productivity was measured by means of two variables — realised and potential productivity.

Both realised (Figure 4) and potential productivity (Figure 5) are highest for elative, but the distinction between internal cases (elative, illative and inessive) is not statistically significant. Allative and adessive have lower realised productivity than internal cases (adessive vs inessive:  $\chi^2 = 5.1514$ ,  $p = 0.02$ ), as well as lower potential productivity than internal cases (adessive vs illative:  $\chi^2 = 4.9469$ ,  $p = 0.03$ ).

The distinction between external and internal cases may be due to the fact that Recipients and Experiencers, which occur with about one fifth of the verbs selecting adessive and allative, often mark human participants (see Section 5.2), leading to a higher proportion of personal pronouns among allative and adessive tokens and therefore a lower type and hapax count. Personal pronouns (singular and plural

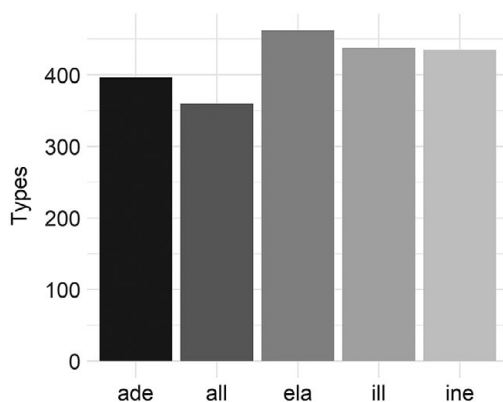


Figure 4

Realised productivity of five spatial cases when marking oblique complements.



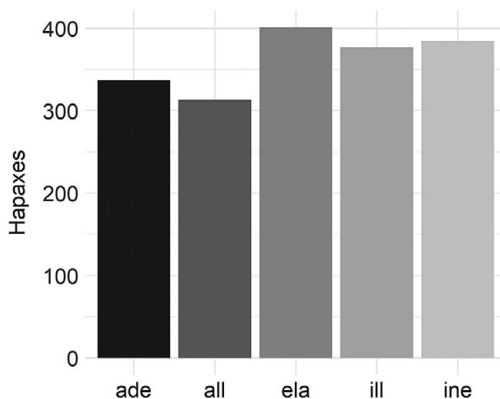


Figure 5

Potential productivity of five spatial cases when marking oblique complements.

forms of 1st, 2nd and 3rd person) accounted for over one quarter (28%) of allative tokens and 18% of adessive tokens in the morphological productivity sample, differing dramatically from the other cases, in which pronouns accounted for 0.8%–8.0% of the complements. These results, taken together, remain somewhat inconclusive for elative and allative, highlighting perhaps a core issue when comparing the morphological productivity of affixes with slightly different syntactic function proportions.

#### 5.4. Syntactic productivity

In the entire sample of oblique complement verbs ( $n = 104$ ), elative and allative case frames are much more frequent than the rest:

As shown in Figure 6, elative is selected by 37% of the verbs in the sample and allative by 38%. Hence, syntactic productivity points to a greater degree of grammaticalisation for elative and allative. This means that these two case frames are more semantically generalised and flexible, making them more easily available when a speaker needs to assign an oblique frame to a novel verb or verb sense. This is most likely also connected to the fact that these two cases have a looser connection to direct semantic effects, as reflected in the roles they mark (Figure 3).

#### 5.5. Summary of results

Research Question 1 asked, Which indirect semantic effects underlie different types of argument structures (oblique and canonical). The study found that these two verb groups are clearly distinct in terms of lexical transitivity. First, K-states, that is states in the sense of Vendler (1957), are significantly more common among oblique complement verbs than among canonical transitive verbs. Second, oblique and

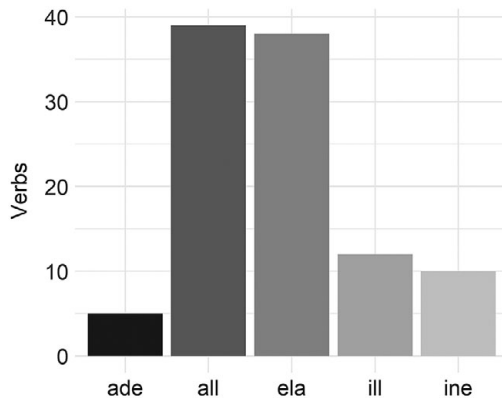


Figure 6  
Syntactic productivity of oblique argument structures.

canonical verbs expressing non-states differ in terms of dynamicity. Hence, on the one hand, states reflecting properties and relationships are likely candidates for non-canonical case frames, but on the other hand, so are verbs depicting non-stative volitional events which do not entail any movement among their participants (ARMUMA ‘fall in love with’). Structure types do not seem to be sensitive to whether or not verbs have natural temporal end points outside K-state proportions.

Research Question 2 asked, how do spatial cases in oblique function differ in terms of direct semantic effects. The study found that elative and allative mark non-spatial roles more frequently than other spatial cases, meaning direct semantic effects are less frequently present for them. This aspect of their meaning was found to be reflected in their individual degrees of grammaticalisation as reflected in syntactic productivity. Morphological productivity did not highlight allative and elative, which may be due to a sampling issue. Namely, this variable was investigated only among case affixes that mark complements, rather than all case affix occurrences in a corpus. This means that cases marking human words have decreased lexical variance, as these words are more frequently pronominalised. Furthermore, looking at productivity in all case instances might be more helpful than just focusing on oblique function, because this is expected to be the most lexically varied function for these cases. In this sense, a distinction of productivity in overall usage is expected to be more drastic than a distinction in oblique use.

## 6. DISCUSSION

The paper set out to investigate two core properties of oblique complement verbs which are thought to differentiate them from canonical argument verbs — indirect and direct semantic effects, using Estonian verbs as data. The results revealed a number of patterns in terms of how verbal semantics affect case selection in verbs.

The results also highlighted the significant variation among oblique cases, some being much more grammaticalised than others, marking highly bleached complements more frequently.

### 6.1. *Indirect semantic effects*

First, the study investigates what are known as indirect semantic effects, referring to the ways in which oblique cases have become linked to certain verbal or complement semantics which are not directly linked to the original meaning of the case. More precisely, we aimed to outline semantic effects related to stativity and its more fine-grained components, lexical dynamicity and lexical aspect. We found that the classic notion of state verbs, that is verbs referring to properties and relationships with distinct morphological, syntactic and semantic behaviour (K-states, e.g. *SÖLTUMA* ‘depend on’, *HOOLIMA* ‘care about’) certainly are more predominant among verbs with non-canonical case frames in Estonian, than among those with canonical frames. In addition, dynamicity was significantly lower among the oblique non-states, compared to the canonically transitive non-states, while the proportion of atelic verbs did not differ between these two groups.

It is not entirely clear whether this means that state verbs and non-dynamic verbs attract non-canonical case frames, or whether it is merely abstract relations in general which attract them, in which case, non-dynamicity would simply be a frequent side product of the property of abstractness in events. In either case, we see an asymmetry between the lack of movement on the part of event participants and the absence of natural time boundaries, as was demonstrated by the fact that the distinction in non-dynamic verb proportions in the two datasets was much more pronounced than the distinction in atelic verb proportions. The fact that non-states in both verb groups have similar telicity proportions also contradicts the hypothesis that aspectual properties of verbs have a strong influence on their syntactic realisation (Tenny 1992). While in this dimension, case does not appear to highlight aspectual distinctions, lexical aspect does play an important role in the differential (canonical) marking of subjects and objects in Estonian. All in all, Estonian oblique complement verbs show strong indirect semantic effects along the dimension of stativity, in terms of having a greater proportion of states and non-dynamic verbs.

### 6.2. *Direct semantic effects*

Second, we outlined the effects directly related to case semantics and revealed their link to the degree of grammaticalisation of the individual cases. Out of five Estonian spatial cases, elative and allative can be said to exhibit direct semantic effects least frequently, marking non-spatial roles with the largest number of verbs. As we hypothesised, both cases also stand out as the most grammaticalised in terms of syntactic productivity. Hence, the study showed that the more grammaticalised the case, the less likely it is for its semantics to be reflected in the meaning of either the

verb or the complement. Elative and allative appear to be widely available for verbs marking K-states and other non-dynamic verbs, mostly without making any spatial references.

It might seem unexpected that instead of stative locative cases, such as inessive ‘in’ and adessive ‘on’, it is the directional cases — separative (elative ‘out of’) and lative (allative ‘onto’) — which have become the predominant markers of non-transitive abstract properties, relations and events. In this regard, the explanation may be diachronic. Partitive, one of the direct object cases in Estonian and Finnish, originates from a separative case (‘from’, semantically similar to the elative ‘out of’). Elative and partitive appear to be in free variation in some contexts marking partiality (POOL LEIBA half bread:PART, POOL LEIVAST half bread:ELA ‘half a bread’). The diachronic link between the other separative case, ablative (‘from on top of’) and partitive has also been noted in French, German, Bulgarian, Lezgian, Krongo, Finnish and Basque (Heine & Kuteva 2002).

The diachronic link between allative and dative is also well documented (Heine & Kuteva 2002: 37). As discussed in Section 2.3, Estonian allative (and adessive) markers are regarded as having both spatial and dative functionality, marking both Goal locations (LAUALE table:ALL ‘onto the table’) and Recipients (EMALE mother:ALL ‘to mother’). This, too, clearly involves abstraction from the original spatial meaning, possibly facilitating the change of meaning in the affix, making it a good candidate for marking other abstract complements without spatial semantics (e.g. allative Themes, see Section 5.2).

Overall, it is problematic to postulate that direct semantics constitute a feature of oblique complement marking. These effects were found to characterise some non-canonical complement cases more than others. This means that this variation exists among cases rather than between types of case frames (oblique vs canonical).

### 6.3. *Implications for argumenthood*

Our results have implications for both how case is used for marking various types of verbal dependency and the relationship between oblique and canonically marked complements in general. As was discussed in Section 1, both Nichols (1983) and Blake (2001) observed semantic distinctions between oblique and canonical complement verbs but drew the opposite conclusions from this observation. While Nichols (1983) suggests that in both oblique and canonical complements, case marks syntactic roles of the same type, Blake (2001) regards the presence of additional semantics (both direct and indirect) as evidence of a categorical distinction between the two types in terms of argument status type/quality.

Our results contribute to the discussion in several ways. First, oblique complement verbs are shown to be semantically rather heterogeneous — a property that has allowed different approaches to use this set of verbs to argue for opposite sides. Even though our dataset was mostly non-dynamic, this does not describe all the verbs in it. Furthermore, this property is certainly not reserved for only oblique

complement verbs, but it is also shared by a number of verbs in the canonical group. Hence, the representation of indirect semantic effects does not neatly map on case frame types, meaning it is problematic to assume their extent and nature in a language. This, on the other hand, means that if indirect semantic effects were to be regarded as a basis for argument status quality, this status, too, could not be viewed as a function of case frame type. It follows that they may not serve as a helpful cross-linguistically universal basis for argument status quality or type.

As for the direct semantic effects, an even more varied picture emerges. Such effects were found to describe only a minority of oblique verbs in our Estonian verb sample, with a sharp distinction between individual oblique cases. Hence, direct semantic effects seem to constitute a property of individual cases rather than oblique arguments. Indirect effects are, therefore, more helpful than direct effects in distinguishing between oblique and canonically marking verbs, but neither describe case frame properties in a clear-cut manner. The links between case frames, direct and indirect semantic effects and argument status quality cannot, therefore, be expressed in any simple manner. The topic needs further research, for instance, by using experimental paradigms where argument status quality is assessed by behavioural means.

Second, our results support a gradient rather than binary approach to argument status (see also, Koenig et al. 2003, Rissman et al. 2015). Morphological case appears to constitute a weak indicator of argumenthood, in that it is a weak or approximative indicator of semantic roles. Entirely bleached roles might occur with a range of cases — a property that does not derive from the semantic meaning of the case. To make matters more complicated, roles themselves appear to be weak indicators of argument status. Studies focusing on particular semantic roles have demonstrated variance in argumenthood judgments for individual roles. For instance, the argumenthood of Instruments varies based on verbal semantics (Koenig et al. 2003, Rissman et al. 2015). Hence, variance in the argumenthood connection quality between verbs and a few other particular constituents is an interesting field which would benefit from a more pluralistic factor-based approach, rather than the perhaps overly simplistic morpho-syntactic definition of argument status.

#### 6.4. *Implications for transitivity*

While oblique case usage might not be a straightforward indicator of the quality of argument status in constituents, our results regarding K-states and non-dynamicity confirm that oblique case tends to co-occur with decreased transitivity. This study used the concept of stativity [Section \(2.1.1\)](#) to assess the role of transitivity, treating it as a similar fundamental expression of this scale. Indeed, the oblique verb group included a much higher proportion of verbs with virtually non-existent lexical transitivity but also verbs with reduced transitivity along the dimension of motion. Given that dynamicity and lexical aspect were shown to play different roles in reducing transitivity, our results suggest that all transitivity factors may not play an equally significant role in case usage and the morphosyntactic organisation of

events. This also highlights the link between argument status and transitivity, suggesting that they are distinct concepts and ought to be treated as such.

Future studies will need to investigate the cross-linguistic role of state categories, dynamicity and aspect in morphosyntax. They will hopefully strive to outline the components of transitivity that are more cross-linguistically predominant among the predictors of case frame types. In addition, diachronic approaches are needed to investigate the tendency of separative semantics grammaticalising into parts of grammar, especially in terms of the bias for expressing Goals over Sources.

#### ACKNOWLEDGMENTS

The authors of this work are grateful to the anonymous reviewers of this paper for their helpful feedback and comments. Some of the results in this article are also included in the main author's doctoral thesis ('In any case? Estonian spatial cases as argument markers', University of Tartu, 2022).

#### REFERENCES

- Ackema, Peter. 2015. Arguments and adjuncts. In Tibor Kiss & Artemis Alexiadou (eds.), *Syntax theory and analysis: An international handbook*, vol 1, 246–274. Berlin: De Gruyter Mouton.
- Aigro, Mari. 2020. Polar question particles and their sources: A semantic approach to grammaticalisation. *Transactions of the Philological Society* 118.2, 237–253.
- Aigro, Mari. 2022. Estonian cases as argument markers. doi:10.17605/OSF.IO/CQRP4. <https://osf.io/cqrp4/>.
- Arka, I Wayan. 2014. Locatives and the argument-adjunct distinction in Balinese. *Linguistic Discovery* 12.2, 56–84.
- Baayen, Harald. 1993. On frequency, transparency and productivity. In Geert E. Booij & Jaap van Marle (eds.), *Yearbook of morphology 1992*, 181–208. Dordrecht: Kluwer Academic Publishers.
- Baayen, Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook*, vol 2, 900–919. Berlin: Walter de Gruyter.
- Barbu, Roxana-Maria & Ida Toivonen. 2016. Event participants and linguistic arguments. In A. Papafragou, Daniel J. Grodner, D. Mirman & J. Trueswell (eds.), *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, 1961–1966.
- Barðdal, Johanna. 2006. Predicting the productivity of argument structure constructions. In Zhenya Antić, Charles B. Chang, Emily Cibelli, Jisup Hong, Michael J. Houser, Clare S. Sandy, Maziar Toosarvandani & Yao Yao (eds.), *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, vol. 32, 467–478. Berkeley, CA: BLS and LSA.
- Barðdal, Johanna. 2008. *Productivity. Evidence from case and argument structure in Icelandic*. Amsterdam and Philadelphia, PA: John Benjamins.
- Blake, Barry J. 2001. *Case*. Cambridge: Cambridge University Press.
- Butt, Miriam. 2006. *Theories of case*. Cambridge: Cambridge University Press.
- Bybee, Joan. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam and Philadelphia, PA: John Benjamins.
- Davidson, Donald. 1967. The logical form of action sentences. In Nicholas Rescher (ed.), *The logic of decision and action*, 81–95. Pittsburgh, PA: University of Pittsburgh Press.
- de Hoop, Helen & Andrej Malchukov. 2007. On fluid differential case marking: A bidirectional OT approach. *Lingua* 117.9, 1636–1656.
- de Hoop, Helen & Peter de Swart. 2009. Cross-linguistic variation in differential subject marking. In Helen de Hoop & Peter de Swart (eds.), *Differential subject marking*, 1–16. New York: Springer.
- Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67.3, 547–619.
- Erelt, Mati. 1989. *Eesti lauseliikmeid (Grammatical relations in Estonian)*. Tallinn: Eesti NSV Teaduste Akadeemia.

- Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael & Silvi Vare. 1993. *Eesti keele grammatika II*. Tallinn: Eesti Teaduste Akadeemia.
- Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael & Silvi Vare. 1995. *Eesti keele grammatika I*. Tallinn: Eesti Teaduste Akadeemia.
- Erelt, Mati. 2003. Süntaks (Syntax). In *Emakeele Seltsi Aastaraamat*, 48, 63–97.
- Erelt, Mati. 2004. Lauseliigenduseprobleeme eesti grammatikas (Issues with parts of speech in Estonian grammar). In Liina Lindström (ed.), *Lauseliikmeist eesti keeles*, 7–16. Tartu: Tartu University Press.
- Erelt, Mati & Helle Metslang. 2006. Estonian clause patterns – From Finno-Ugric to Standard Average European. *Linguistica Uralica* 42.4, 254–266.
- Fábregas, Antonio & Rafael Marín. 2012. State nouns are Kimian states. In Irene Franco, Sara Lusini & Andres Saab (eds.), *Romance languages and linguistic theory 2010: Selected papers from 'going romance' leiden 2010*, 41–64. Amsterdam and Philadelphia, PA: John Benjamins.
- Forker, Diana. 2014. A canonical approach to the argument/adjunct distinction. *Linguistic Discovery* 12.2. doi:10.1349/PS1.1537-0852.A.444.
- Haspelmath, Martin. 2001. Non-canonical marking of core arguments in European languages. In Alexandra Y. Aikhenvald, R. M. W. Dixon & Masayuki Onishi (eds.), *Non-canonically marked subjects and objects*. Amsterdam: John Benjamins Publishing Company.
- Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Hiietam, Katrin. 2003. *Definiteness and grammatical relations in Estonian*. Ph.D dissertation, University of Manchester.
- Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56.2, 251–299.
- Hopper, Paul J. & Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.
- Kasik, Reet. 2015. *Sõnamoodustus (Word formation)*. Tartu: Tartu University Press.
- Kehayov, Petar & Reeli Torn-Leesik. 2009. Modal verbs in Balto-Finnic. In Björn Hansen & Ferdinand de Haan (eds.), *Modals in the languages of Europe*, 363–401. Berlin and New York: Mouton de Gruyter.
- Kettunen, Lauri. 1924. *Lauseliikmed eesti keeles (Parts of speech in Estonian)*. Tartu: Eesti Kirjanduse Seltsi kirjastus.
- Kim, Jaegwon. 1969. Events and their descriptions: Some considerations. In Nicholas Rescher (ed.), *Essays in honor of Carl G. Hempel*, 198–215. New York: Springer.
- Kittilä, Seppo. 2008. Case and the typology of transitivity. In Andrej Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 356–369. Oxford: Oxford University Press.
- Klaas, Birute. 1988. Indirektne objekt (Indirect object). *Keel ja Kirjandus* 1, 1–8.
- Klaşlan, Yılmaz & Gurkan Tuna. 2015. Linking FrameNet and natural languages via thematic role structures. *International Journal of Languages, Literature and Linguistics* 1.3, 158–163.
- Koenig, Jean Pierre, Gail Mauner & Breton Bienvenue. 2003. Arguments for adjuncts. *Cognition* 89.2, 67–103.
- Kratzer, Angelika. 2004. Telicity and the meaning of objective case. In Jacqueline Guéron & Jacqueline Lecarme (eds.), *The syntax of time*. Cambridge, MA: MIT Press.
- Kure, Kristjan. 1959. Eesti keele lauseliigenduse alustest. *Keel ja Kirjandus* 1, 40–49.
- Lakoff, George. 1966. Stative adjectives and verbs in English. In Anthony G. Oettinger (ed.), *Mathematical linguistics and automatic translation*. Cambridge, MA: Aiken Computation Laboratory, Harvard University.
- Lestrade, Sander. 2010. *The space of case*. Ph.d dissertation, Radboud University.
- Lindström, Liina. 2017. Partitive subjects in Estonian dialects. *Journal of Estonian and Finno-Ugric Linguistics* 8.2, 191.
- Lindström, Liina & Ilona Tragel. 2010. The possessive perfect construction in Estonian. *Folia Linguistica* 44.2, 371–400.
- Lindström, Liina & Virve Anneli Vihman. 2017. Who needs it? Variation in experiencer marking in Estonian 'need'- constructions. *Journal of Linguistics* 53.4, 789–822.
- Maienborn, Claudia. 2003. Event-internal modifiers: Semantic underspecification and conceptual interpretation. In Ewald Lang, Claudia Maienborn & Cathrine Fabricius-Hansen (eds.), *Modifying adjuncts*, 475–509. Berlin: Mouton De Gruyter.
- Maienborn, Claudia. 2008. On Davidsonian and Kimian states. In Ileana Comorovski & Klaus von Heusinger (eds.), *Existence: Semantics and syntax*, vol. 84, 107–130. Dordrecht: Springer.



- Maienborn, Claudia. 2019. Events and states. In Robert Truswell (ed.), *The Oxford handbook of event structure* January, 49–89. Oxford: Oxford University Press.
- Malchukov, Andrej & Peter De Swart. 2008. Differential case marking and actancy variations. In Andrej Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*. Oxford: Oxford University Press.
- Malchukov, Andrej L. 2005. Case pattern splits, verb types and construction competition. In Mengistu Amberber & Helen de Hoop (eds.), *Competition and variation in natural languages*, 73–117. Elsevier Science.
- Matsumura, Kazuto. 1994. Is the Estonian adessive really a local case? *Journal of Asian and African Studies* 46–47, 223–235.
- Metslang, Helena. 2013. *Grammatical relations in Estonian: Subject, object and beyond*. Ph.D dissertation, University of Tartu.
- Mihkla, Karl, Lehte Rannut, Elli Riikoja & Aino Admann. 1974. *Eesti keele lauseõpetuse põhijooned I (Foundations of Estonian Syntax I)*. Tallinn: Valgus.
- Miljan, Merilin. 2009. *Grammatical case in Estonian*. Ph.d dissertation, University of Edinburgh.
- Nichols, Johanna. 1983. On direct and oblique cases. In *Proceedings of the Ninth Annual Meeting of the Berkeley Linguistics Society (1983)*, 170–192. Berkeley, CA.
- Nichols, Johanna. 1984. Direct and oblique objects in Chechen-Ingush and Russian. In Frans Plank (ed.), *Objects. Towards a theory of grammatical relations*, 183–209. London: Academic Press Inc.
- Nurka, Maiu. 2014. Eesti keele elatiivi semantika kognitiivse keeleteaduse vaatenurgast (Semantics of Estonian elative from a cognitive linguistics perspective). Master thesis, University of Tartu.
- Onishi, Masayuki. 2001. Non-canonically marked subjects and objects: Parameters and properties. In Alexandra Y. Aikhenvald, R.M.W Dixon & Masayuki Onishi (eds.), *Non-canonical marking of subjects and objects*, 1–53. Amsterdam and Philadelphia, PA: John Benjamins.
- Pilvik, Maarja-Liisa. 2019. Assessing the productivity of the Estonian deverbal suffix -mine in five registers of Estonian. *SKY Journal of Linguistics* 32, 75–103.
- Primus, Beatrice. 1999. *Cases and thematic roles*. Berlin: De Gruyter.
- Proos, Mariann. 2021. *Meaning and usage of Estonian experience perception verbs*. Ph.D dissertation, University of Tartu.
- Rajandi, Henno & Helle Metslang. 1979. *Määramata ja määratud objekt (Definite and indefinite objects)*. Tallinn: Valgus.
- Rätsep, Huno. 1978. *Eesti keele lihtlause tüübid (Clause types in Estonian)*. Tallinn: Valgus.
- Rätsep, Huno. 1979. *Eesti keele ajalooline morfoloogia II (Historical morphology of Estonian II)*. Tartu: Tartu National University.
- Rissman, Lilia & Asifa Majid. 2019. Thematic roles: Core knowledge or linguistic construct? *Psychonomic Bulletin and Review* 26.6, 1850–1869.
- Rissman, Lilia, Kyle Rawlins & Barbara Landau. 2015. Using instruments to understand argument structure: Evidence for gradient representation. *Cognition* 142, 266–290.
- Rothmayr, Antonia. 2009. *The structure of stative verbs*. Amsterdam and Philadelphia, PA: John Benjamins.
- Saluveer, Viktor. 1958. Sise- ja väliskohakäänete tähenduslikud funktsioonid eesti keeles. Master thesis, Tartu National University.
- Tenny, Carol. 1992. The aspectual interface hypothesis. In Ivan. A. Sag & Anna Szabolcsi (eds.), *Lexical matters*, 1–27. Stanford: CSLI Publications.
- Tsunoda, Tasaku. 1981. Split case-marking patterns in verb-types and tense/aspect/mood. *Linguistics* 19, 389–438.
- Valgma, Johannes & Nikolai Rimmel. 1968. *Eesti keele grammatika (Estonian grammar)*. Tallinn: Valgus.
- Veismann, Ann, Mati Erelt & Helle Metslang. 2017. Määrus (Adverbial). In Mati Erelt & Helle Metslang (eds.), *Eesti keele süntaks (Estonian syntax)*, 300–375. Tartu: Tartu University Press.
- Vendler, Zeno. 1957. Verbs and times. *The Philosophical Review* 66.2, 143–160.
- Verkuyl, Henk J. 1999. *Aspectual issues. Studies on time and quantity*. Chicago, IL: University of Chicago Press.
- Viht, Annika & Külli Habicht. 2019. *Eesti keele sõnamuutmine (Estonian inflectional morphology)*. Tartu: Tartu University Press.

OBLIQUE COMPLEMENTS IN ESTONIAN: A CORPUS PERSPECTIVE

- Zaenen, Annie & Joan Maling. 1984. Unaccusative, passive, and quirky case. In Joan Maling & Annie Zaenen (eds.), *Modern Icelandic syntax*, 137–152. Leiden: Brill.
- Zeldes, Amir. 2012. *Productivity in argument selection: From morphology to syntax*. Berlin and Boston, MA: De Gruyter Mouton.

*Authors' addresses:* (Aigro)  
*University of Tartu, Jakobi 2-430, 51005, Tartu, Estonia*  
*mari.aigro@ut.ee*

(Vihman)  
*University of Tartu, Jakobi 2-416, 51005, Tartu, Estonia.*