Who needs it?
Variation in experiencer marking in
Estonian 'need'-constructions[1]

LIINA LINDSTRÖM & VIRVE-ANNELI VIHMAN

University of Tartu

In this paper, we tackle the twin issues of obligatoriness of semantic arguments and variation in their expression through a study of Estonian constructions denoting need. The variation under investigation consists in the choice of case-marking, between adessive and allative case, as well as the option to omit the oblique argument. We extracted and coded 'need'-constructions from spoken and written corpora and used non-parametric classification methods for analysis. We found high rates of oblique experiencer omission in these constructions (nearly 60% across corpora). The most important predictors of overt expression of the experiencer in our models were participant-internal modality and the presence of nominal complements, meaning that both semantic and syntactic factors are relevant. The choice between two overt cases is affected by person, complement type, and referential distance. Topical experiencer arguments do not show the subject-like tendency to be omitted more often, but they are more likely to be marked with adessive case, suggesting that adessive is more grammaticalised as a structural, non-nominative, argument-marking case than the more semantic allative case. Our findings show that oblique, semantic arguments may be frequently omitted, and both semantic and syntactic factors may affect variation in case-marking.

KEYWORDS: argument omission, experiencer, modality, variation in case-marking

1. INTRODUCTION

Non-prototypical argument marking can reveal aspects of the case-marking system of a particular language, but can also speak to broader questions, such as which arguments are prone to atypical marking, why, and what generalisations

---

can be made about alternative case-marking cross-linguistically. Experiencers are excellent candidates for atypical marking, as they are intimately linked to the core semantics of a predicate, and typically human, but unagentive, hence atypical in either agent or patient roles. Options for expressing the experiencer argument vary greatly across languages and across constructions, in terms of case-marking, word order, and alignments with prototypical arguments. This variation has been the subject of a good deal of fruitful research, both cross-linguistically (see Croft 1993, Bossong 1998, Haspelmath 2001, Bickel 2004, de Hoop & de Swart 2009, among many others) and language-internally (for Estonian, see e.g. Erelt & Metslang 2008, Lindström 2013). The focus of this paper is language-internal variation in expression of the experiencer argument in two related constructions expressing need in Estonian, one with a nominal complement, as shown in (1a), the other with an infinitival complement, as in (1b).

(1) (a) Mu-l/Mu-lle **on**   **vaja** uut   arvuti-t.
          I-ADE/I-ALL be.PRS.3SG need new.PRT computer-PRT

          'I need a new computer.'

     (b) Mu-l **on**   **vaja** koju   minna.
          I-ADE be.PRS.3SG need home.ILL go.INF

          'I need to go to home.'

The two arguments in experiencer constructions are less distinct in terms of control and affectedness than in canonical transitives (Næss 2007: 190), and this difference has grammatical effects, including variability in word order and variation in the case-marking of the experiencer argument. Thus, variation in experiencer marking is often associated with semantic issues and connected to predicate semantics (see e.g. Croft 1993).

The MODAL EXPERIENCER ARGUMENT often patterns with 'non-canonically marked subjects', or core arguments which diverge from prototypical subjects in certain morphosyntactic coding properties yet also exhibit behaviour similar to prototypical subjects (see Onishi 2001, Narrog 2010). Modal constructions have been treated less often from this perspective. The form of expression of the modal experiencer argument may be related to modal semantics, syntactic context or even discourse-level pragmatics. Additionally, the question arises whether these should be considered syntactic arguments of the predicate. This paper investigates the experiencer argument and factors affecting variation in its expression in the Estonian 'NEED'-CONSTRUCTIONS. It has been shown that some of the variation we see in these constructions can be traced diachronically to contact with neighbouring languages (Lindström, Uiboaed & Vihman 2014). Here, we ask how to account for synchronic variation in the marking of the experiencer argument and whether this can be illuminating for cross-linguistic approaches to omission and argument behaviour. In a typological study of the associations between modality and non-canonical case-marking, Heiko Narrog finds that only 'about a quarter of the languages in the sample . . . had modal

constructions involving either voice or non-canonical marking', adding that 'there is an apparent association of the necessitive with non-canonical marking' (Narrog 2010: 82). The functions typically served in European languages by the dative case (a common contender for marking non-canonical, subject-like arguments) are shared by the external locative ADESSIVE and ALLATIVE cases in Estonian in e.g. possessive, psychological and cognition predicates. Connections between dative case and allative case have been noted elsewhere as well (Blansitt 1988; Næss 2007; Creissels 2009: 621). Among European languages, non-canonical subjects are used in modal constructions primarily in East Slavonic and Baltic languages (Hansen 2014), but this pattern is also widespread in Balto-Finnic languages, including Estonian (Kehayov & Torn-Leesik 2009). Hence, this may be an areal phenomenon.

In addition to the oblique, adessive/allative marking, there is also an option to omit the experiencer in the constructions under discussion. Typologically, null reference is rare in the oblique position (Siewierska 2003); however, it is frequently attested in Estonian experiencer constructions. Omission of the modal argument is also possible in some other languages that make use of non-canonical marking of modal experiencer subjects, such as Baltic (Holvoet 2007) and Slavic languages (Hansen 2014), but what motivates experiencer omission is not clear.

Both the oblique case-marking of the experiencer and its propensity to be omitted lead to the question of whether it really is a (non-prototypical) argument, or rather ought to be given adjunct status. Semantic cases in Finnic are generally seen to encode adverbial constituents with semantics similar to adpositional phrases; their behaviour in argument-like contexts has been less thoroughly examined in the literature, and so their interpretation and treatment in these contexts remains open for discussion (EKG II: 61). However, the experiencer is clearly a semantic argument, denoting a core participant in the event: this is based on intuition, and corresponds to Needham & Toivonen's (2011) CORE PARTICIPANTS test. Additionally, the experiencer also passes the VERB SPECIFICITY test of Koenig, Mauner & Bienvenue (2003), being specific to the individual event type and lexically required to bear certain properties. However, 'despite its foundational importance within syntactic theory, the argument/adjunct distinction has never been very well defined and there exist gray areas in the taxonomy' (Tutunjian & Boland 2008: 632). Koenig et al. (2003: 68) note that 'while most linguists agree that the distinction between arguments and adjuncts is real, no consensus currently exists as to its basis, the boundary between the two classes, or its role in grammar'. Data from languages like Estonian, and the atypical experiencer arguments investigated here, cast further doubt on the reliability and discreteness of the distinction (see also Creissels 2014, who argues, on the basis of a crosslinguistic analysis of beneficiaries, that semantic argumenthood should be seen as a scalar concept). The use of oblique cases to represent arguments required by the logical structure of an event, and the referential interpretation present even when these arguments are omitted, presents a confound in the ability to distinguish between arguments and adjuncts.

We investigate two constructions with many overlapping properties and which exhibit similar variability in expressing the experiencer argument. What governs the choice between the three options for experiencer (non-)expression – adessive, allative or zero – is not clear from earlier studies. Our goal is to clarify what factors influence experiencer coding, and whether they are primarily semantic or syntactic. In order to eliminate distractors, we divide the study into two tasks. First, we look at conditions affecting the choice between explicit and implicit experiencers. Omission of both subject and object arguments is common in Estonian, and thus far no dedicated studies have investigated what motivates the omission of oblique arguments, nor whether a principled distinction can be made between oblique argument omission and adjunct omission. Second, we narrow in on the overt experiencers and look at the factors affecting the choice between adessive and allative case-marking.

We aim to reveal how much of the variation can be explained by semantics, and we find that both semantic and syntactic factors play a role in the variation. The overt expression or omission of the experiencer turns out to be closely connected to clausal semantics, and whether the need or necessity can be ascribed to participant-internal or participant-external modality (see van der Auwera & Plungian 1998). Among overt arguments, however, the choice of case is affected by person, complement type and accessibility of the referent. Hence, the presence or absence of the experiencer argument is semantically motivated, whereas both syntactic and discourse-pragmatic factors more often related to subject expression are at play in the choice of case-marking. The motivation for argument encoding may vary greatly, depending on the language, construction, and context.

We first give an overview of the Estonian 'need'-constructions under investigation, the variation in experiencer marking and its consequences for analysis, as well as subject ellipsis. In Section 3, we describe the methodology used in our study. In Sections 4 and 5, we report on our results and analysis, and discuss their implications.

2. CHARACTERISTICS OF THE CONSTRUCTIONS

2.1 *Estonian 'need'-constructions*

Experiencer arguments occur with a range of case-marking patterns in Estonian, from nominative subject-marking to partitive, adessive and allative case with various experiencer and cognition predicates. The 'need'-constructions investigated here can include either adessive or allative experiencers. The predicate in both constructions involves a copula verb (*olema* 'be'; *on* 'is' in examples (1a–b) above and (2a–c) below) and the adverb *tarvis* or *vaja*, expressing need, necessity or obligation; for present purposes, we consider *tarvis* and *vaja* to be interchangeable (the examples in (1) and (2) use *vaja*). In Estonian grammars (EKG II, Erelt 2013) the modal predicates *tarvis olema* and *vaja olema* are listed as synonymous, expressing the same meanings ('to need') and functions.

792

In addition to the synonymous adverbs of necessity, every element of the constructions can vary: the experiencer is marked with adessive or allative case, and can be omitted; the copula can be replaced by *tulema* 'come' or *minema* 'go', or can be omitted; and the verbal complement also exhibits variation, as detailed below. Hence, these constructions are characterised by great variability across uses, and they have been found to reflect differences according to dialect, region and genre (Lindström et al. 2014).

Among these variables, we distinguish two basic constructions according to their complements, shown in (2), as these represent substantive differences in the form–meaning pairing (SPO indicates data from the Corpus of Spoken Estonian and WRI indicates examples from the Corpus of Written Estonian).

(2) (a) mu-l on   **kassetti**   **vaja**.
I-ADE be.3SG cassette.PRT need

'I need a cassette.'

(b) mu-l on   **vaja piljardilauda**   **liiguta-da**.
I-ADE be.3SG need billiard.table.PRT move-INF
(http://www.sgbilliard.ee/kkk/kysimused-laudade-kohta)
'I need to move a/the billiard table.'

(c) kõigepealt on   **vaja**, **et** **perearst**   **käi-ks**.   (SPO)
first   be.3SG need that family.doctor visit-CND

'first the doctor has to pay a visit.'
(lit.: 'first it is necessary that the doctor visit.')

(d) ... siis ei õppi-nud sa rohkem, **kui eksame-i-ks vaja**? (SPO)
then not study-APP you more   than exam-PL-TRL need

'then you didn't study any more than [you] needed [to] for the exams?'

The construction may be used either with a nominal complement (marking need for something), as in (2a), an infinitival complement (marking necessity or obligation to do something), as in (2b), or a finite clausal complement, as in (2c). The construction with a clausal complement is rare, at least in our data. In naturally occurring data, the construction can also occur without any complement, in which case the object of need must be inferred from the (discourse) context which, as in (2d), exhibits both elided experiencer and complement. As each of these related constructions is included in our analysis, we use the term 'NEED'-CONSTRUCTIONS to cover all of them, and the term (MODAL) EXPERIENCER for the semantic argument expressing the experiencer.

The constructions we have chosen to investigate are distinguished only by the form of the predicate complement. In each of these constructions, the finite verb is in default third person singular form and does not show agreement with the experiencer argument. Hence, experiencer omission in these constructions cannot be explained by redundant person marking on the verb, as is often the case with omitted nominative subjects; in these cases with experiencer ellipsis, the

experiencer referent must be resolved through contextual or discourse-pragmatic cues.

While the construction with a nominal complement (as in (2a)) expresses a relation of need for, or lack of, the partitive NP referent – and can therefore be classified as 'premodal' usage – the infinitival construction is more grammaticalised as a true modal construction, marking necessity or deontic modality, with the main predicate, *tarvis/vaja olema*, and the infinitival complement acting together as modal predication. With the infinitival complement, the construction expresses either participant-internal dynamic necessity, as in example (3), participant-external (non-deontic) necessity (referring to circumstances external to the participant), as in (4), or deontic necessity, as in (5).

(3) Mu-l **on** **vaja** söö-ma **minna**.
I-ADE be.3SG need eat-SUP go.INF

'I need to go eat.'

(Erelt 2013: 110)

(4) Ne-i-l **ol-i** **vaja lahku-da**, sest tuul hakka-s
they-PL-ADE be-PST.3SG need leave-INF because wind start-PST.3SG
tõus-ma.
rise-SUP

'They had to leave because it was getting windy.'

(Erelt 2013: 113)

(5) Te-i-l **on** **vaja** homme ise koha-l **olla**.
you-PL-ADE be.3SG need tomorrow self place-ADE be.INF

'You have to be there yourself tomorrow.'

(Erelt 2013: 114)

2.2 *Argument ellipsis*

Estonian is considered a partial pro-drop language, allowing optional ellipsis of subjects. According to Duvallon & Chalvin (2004: 272), 18% of first-person singular forms and nearly half of all second-person singular verbs in their analysis of spoken data have zero subjects. Dialect data analysed by Lindström et al. (2009) exhibited omission of first-person singular subject pronouns in 11–54% of examples, showing great variability across dialects. In the case of nominative subjects with verb agreement, the verb inflection provides unambiguous information for first and second-person referents. Third-person referents require more context to support reference resolution, and therefore ellipsis is used mainly with highly accessible (given) referents, for example in narrative passages (Hint 2015). Third-person referents may also be less frequently dropped because clauses with third-person singular verb inflection

794

can give rise to a generic, 'zero-person' reading (Kaiser & Vihman 2007, Jokela 2012).

With constructions such as those in our study, which lack verb agreement, the generic reading may be especially prominent, as in (6), which can be read out of context as a generic clause, whereas the narrative context of the example confirms that the clause attributes the need to a referent named in the preceding sentence.

(6) Kui vähe **ol-i** **tarvis** õnne-ks ja rahu-ks! (WRI)
   how little be-PST.3SG need happiness-TRL and peace-TRL

   'How little was needed for happiness and peace!'

Experiencer omission is also found in modal contexts where the experiencer is not general but specific, typically referring to the speaker or addressee, although verb agreement does not support reference resolution here as it does in canonical clauses:

(7) Ise ol-i-me süüdi, pole-ks **tarvis** ol-nud minna
   self be-PST-1PL guilty be-CND.NEG need be-APP go.INF
   regionaalliini-le, seal raputa-b rohkem. (WRI)
   regional.line-ALL there shake-PRS.3SG more

   'We had only ourselves to blame, we shouldn't have gone on the local bus service, where it's bumpier.'

Example (8), from the Corpus of Written Estonian, makes explicit use of the ambiguity between generic and referential readings of the zero experiencer, as can be seen in the text following the direct quote, in (8b).

(8) (a) "Küll ma tule-n, kui tunne-n, et **on** **tõesti vaja**,"
      PRTCL I come-1SG if feel-1SG that be.3SG really need
      ütle-s ta Mauno-le.
      say-PST.3SG he Mauno-ALL

      "'Sure I'll come if I feel that it's really necessary," he said to Mauno.'

   (b) Selgita-ma-ta täpsemalt, kas vaja ta-lle enese-le
      clarify-SUP-ABE precisely whether necessary he-ALL self-ALL
      või palju-de-le teis-te-le. (WRI)
      or many-PL-ALL other-PL-ALL

      'Without clarifying whether he meant NECESSARY FOR HIMSELF OR MANY OTHERS.'

The omission of experiencer arguments in contexts which do not unambiguously support structural ellipsis (i.e. without overt person marking) has received insufficient attention in the literature, and may affect the way we think of the binary distinction between arguments and adjuncts. Björn Hansen notes that omission of non-canonical modal subject arguments is common in East Slavonic, Baltic and Balto-Finnic languages (Hansen 2014 for Finnic languages; see also Kehayov &

795

Torn-Leesik 2009), and, as we demonstrate, Estonian is no exception. However, the conditions, functions and consequences of this argument omission have not been elucidated.

We address this gap by investigating what conditions favour the omission of the modal experiencer. One important question to ask here is to what extent the omission of modal experiencer arguments looks like subject ellipsis. The modal experiencer has many similarities with canonical subjects, such as a preference for human referents and preverbal pronouns. Hence, we ask whether omission of the subject-like experiencer occurs under similar conditions as omission of canonical, nominative subjects, and if not, what factors motivate modal experiencer drop; further, we may ask whether the conditions of omission are more similar to oblique adjuncts or arguments.

Earlier studies on Estonian do not address these questions directly, but some indications can be found. Penjam (2006) finds that, in written Estonian, the adessive modal experiencer argument occurs overtly in only one fifth of occurrences of the modal construction with the predicate *tulema* 'come; have to, need to'. Omission of the experiencer often occurs when referring to speech act participants, especially the speaker. This is particularly common in contexts of negative politeness (e.g. internet fora) in which open reference to the speaker or addressee is often avoided (Lindström 2010, also see Zinken & Ogiermann 2011 on a similar strategy in 'need'-clauses in Polish). However, the tendency to omit the experiencer is not restricted to contexts of negative politeness or particular constructions, but is widespread in most of the experiencer constructions.

Zero anaphora (i.e. argument omission) is included in Metslang (2013) as a feature characteristic of subjects in Estonian, albeit a statistical feature, which Metslang considers to be weaker than more categorical behavioural characteristics. She finds argument omission to be most characteristic of the A argument (subject of transitive constructions, omitted in 39% of examples), the single argument (S) in intransitive constructions (30%), and the allative argument in experiencer constructions. According to Metslang, the allative experiencer is omitted in 28% of her examples, though note that no adessive experiencers are included; importantly, in the case of elided arguments, this probably means a conflated result. Metslang reports much smaller rates of omission (2–6%) for other analysed arguments (derived subject, possessor, possessee, stimulus, transitive object; Metslang 2013: 245). Thus, dative-like experiencers tend to be omitted often, on a level similar to canonical nominative subjects and far more than other non-nominative arguments, but less often than nominative S or A arguments.

Our study, then, picks up some of the uninvestigated areas of argument omission, asking how frequently experiencers are omitted in the constructions under investigation, what conditions favour omission, and how similar this is to subject ellipsis. This includes the investigation of whether referential distance plays a role, as well as what semantic and syntactic factors most affect overt experiencer expression. As syntactic factors affecting experiencer omission are

796

not readily apparent, our first hypothesis (stated explicitly in Section 2.5 below) postulates that experiencer omission is connected to a semantic factor, namely type of modality. We discuss this in Section 2.4, below.

2.3 *Adessive and allative case-marking of experiencers*

Analyses of non-canonically marked subject-like arguments show that, typically, where case-marking exhibits variation, the choice of case depends on the predicate; in other words, predicates with particular semantics require the higher (non-canonically marked) argument to take a particular form. In this light, the 'need'-constructions investigated in this paper are remarkable in that they form a distinct semantic class but nevertheless exhibit case alternation between adessive and allative arguments.

In Estonian grammars, the category of subject is defined mainly by coding properties: nominative case, predicate agreement, and typically clause-initial position. Certain sentence types, however, include non-nominative arguments which bear behavioural and/or coding properties similar to subjects. Partitive, adessive and allative case can mark arguments which have acquired subject properties to varying degrees, e.g. adessive possessors (Erelt & Metslang 2006), adessive agents in passive clauses (known as the 'possessive perfect' construction and exemplified below in example (10), Lindström & Tragel 2010), and partitive experiencers in the experiencer-object construction (Lindström 2013). These constructions have low transitivity but typically human experiencer or agent referents; they show no predicate agreement with the subject-like argument. No comprehensive overview exists of non-canonical subject constructions in Estonian.

Traditional grammars usually treat adessive and allative (dative-like) arguments as adverbials, not as core predicate arguments. Metslang (2013) measures the subjecthood of subject-like arguments in comparison with prototypical, nominative subjects. She concludes that oblique topical arguments in marked clauses (possessor, experiencer, source) are less subject-like, as they pass a smaller number of subject tests. The tests they pass are mostly statistical tendencies (preverbal word order, zero anaphora), whereas they fail most of the behavioural tests (Metslang 2013: 284). However, Metslang does not analyse adessive experiencers, which occur with a broader range of predicates than the allatives included in her study. Lestrade (2010: Chapter 5), discussing the use of spatial cases to express structural relations, makes the claim that the use of spatial case is semantically motivated, forcing the suspension of some inappropriate inferences in interpreting a human or animate argument referent, such as animacy and topicality. We would like to claim that in Estonian, this may have motivated the use of both adessive and allative for marking experiencer arguments, but that the adessive is so commonly used in non-canonical argument marking in modern Standard Estonian that it no longer bears this semantic markedness. Both continue to be used as locative case-markers as well.

The core function of the local adessive and allative cases is to express spatial relations: adessive expresses the static relation 'on (top of)', and allative expresses the directional 'onto' (see e.g. Klavan, Kesküla & Ojava 2011). In addition to locative semantics, adessive case is used with human participants to mark a variety of relations, including possessor, as in (9) (see Erelt & Metslang 2006), external possessor, 'possessive perfect' agent, as in (10) (Lindström & Tragel 2010, Lindström 2015), experiencer, as in (11), and certain modal relations.

(9) Mu-l   on   uus   auto.   (adessive possessor)
     I-ADE be.3SG new.NOM car.NOM

     'I have a new car.'

(10) Mu-l   on   auto   pes-tud.   (adessive agent in possessive perfect)
     I-ADE be.3SG car.NOM wash-PPP

     'My car is washed./ I've got my car washed.'

(11) Mu-l   on   tema-st kahju.   (adessive experiencer)
     I-ADE be.3SG s/he-ELA sorry

     'I feel sorry for him/her.'

Allative case is typically used with human participants to mark recipient or beneficiary roles, as shown in (12) below. In some experiencer constructions, allative can also be used to mark the experiencer argument, e.g. with predicates like *meeldima* 'like/please', as in (13), *meelde tulema/meenuma* 'remember', as in (14), or *tunduma* 'seem'.

(12) Ta   and-i-s   mu-lle medali.   (allative recipient)
     s/he.NOM give-PST-3SG I-ALL medal.GEN

     'S/he gave me a medal.'

(13) Mu-lle meeldi-b su   uus   müts.   (allative experiencer)
     I-ALL like-3SG you.GEN new.NOM hat.NOM

     'I like your new hat.'

Most of the constructions above are associated with particular case-marking. However, some experiencer predicates allow variation in case-marking strategies, especially between adessive and allative in many experiencer and modal constructions, as exemplified in (14).

(14) Mu-lle/mu-l tul-i   meel-de üks   lugu.
     I-ALL/I-ADE come-PST-3SG mind-ILL one.NOM story.NOM

     'A story came to my mind (I recalled a story).'

The motivation behind this variation is not immediately apparent, and does not seem to be associated with distinct semantics. Though the apparently free variation in case-marking has not been well studied, the variation itself has long roots. Extensive variation between adessive and allative case has been noted in Old

Written Estonian as early as the 17th century (Ross 1997) and in Estonian dialects (Pajusalu 1958). However, the variability in current standard usage demands an explanation, even if its background lies in dialectal differences.

As shown above, the adessive case has been extended to mark a number of non-spatial relations in a variety of constructions. The result is that purely spatial semantics accounts for a smaller portion of adessive uses, whereas argument structure and constructional semantics play a more important role in interpretation, as the adessive has undergone a degree of grammaticalisation to mark predicate arguments in various constructions. The allative case, on the other hand, is still more strongly associated with a directional meaning ('onto'), while also including the non-spatial semantic extension 'for', allowing benefactive (and malefactive) readings. The allative is more associated with adverbial adjuncts than the adessive, but the contexts of variation between the two occur in argument positions.

Thus, we argue that in Estonian, the better candidate for subject, among non-nominative nominals, is the adessive NP, while allative is more restricted in usage and interpretation. Example (14) above shows that *meelde tulema* 'recall, come to mind' can be used with either an adessive or allative experiencer with a nominal complement. This contrasts with example (15) below, where typically only the adessive experiencer would be used for the same predicate with an infinitival complement, although the allative is not entirely ungrammatical. Here, we see variation in which the adessive appears to be taking over allative functions for marking the experiencer argument.

(15) Mu-l/?mu-lle   tul-i   meel-de enne pimeda-t koju
      1SG-ADE/?I-ALL come-PST.3SG mind-ILL before dark-PRT home.ILL
      minna.
      go.INF
      'I remembered to go home before dark.'

The ability to take an infinitival complement is a step along the grammaticalisation (or 'constructionalisation') chain (the issue is mostly discussed in connection with auxiliarisation, see Bolinger 1980, Heine 1993, Kuteva 2001, among many others). This leads us to postulate our Hypothesis 2, that the construction with the – more grammaticalised – infinitival complement more frequently selects the adessive experiencer argument (see hypotheses in Section 2.5 below).

2.4 *Modality*

The notions of participant-internal and participant-external modality come from the typology of modality in van der Auwera & Plungian (1998), later adjusted by van der Auwera, Kehayov & Vittrant (2009). Participant-internal modality refers to possibility or necessity that is 'internal to a participant engaged in the state of affairs' (van der Auwera, Plungian 1998: 80), i.e. in the case of necessity, it implies that the need originates from within the participant, as in (16a).

799

(16) (a) Ne-i-l   ol-i   tarvis süüa.   (WRI)
          they-PL-ADE   be-PST.3SG   need   eat.INF
          'They needed to eat.'

      (b) kes (.) ee taha-ks (.)   küsimus-te-le   vastami-st (.)
          who   uh want-CND.3SG   question-PL-ALL   answering-PRT (.)
          kus   on   vaja pikalt kirjuta-da.   (SPO)
          where be.3SG need long   write-INF
          'Who would want questions where [you] need to write long answers?'

Participant-external modality involves possibility or necessity ascribed to causes external to the participant, as in (16b), referring to exam questions requiring written answers (for examples from the spoken corpus, the notation (.) indicates a short pause; a number in parentheses in such examples, e.g. (17b, c) below indicates the length of the pause in seconds). Deontic modality is treated as a subtype of participant-external modality: according to van der Auwera & Plungian (1998: 81), it 'identifies the enabling or compelling circumstances external to the participant as some person(s), often the speaker, and/or some social or ethical norm(s) permitting or obliging the participant to engage in the state of affairs'.

The notions of participant-internal and participant-external modality are applied here not only to modal necessive or deontic constructions with infinitival complements, but also to clauses with nominal complements and without complements, usually outside the scope of analyses of modality, as these constructions do not form proper modal verb constructions, but rather 'premodal' usages. In the constructions with *tarvis*/*vaja*, it was possible to make a distinction between internal, (17a), and external, (17b), sources of need. However, we present results both with and without the premodal uses included.

(17) (a) mu-l ole-ks kampsi   vaja.   (SPO)
          I-ADE be-CND sweater.PRT need
          'I need a sweater.'

      (b) A: valan teed sulle. (.)   (SPO)
              'I'll pour you some tea.'

          B: oota m (.) oota ma toon sõela vä. (1.2)
              'Wait. Wait, should I bring you a sieve?'

          A: tea   **kas   se-da   sõela   tarvis** see (0.5)
              know.CONNEG whether this.PRT sieve.PRT need   this
              vist   on (1.2) suure-d tüki-d.
              probably   be.3PL   big-PL   piece-PL
              'Don't know if the sieve is really necessary. I guess it is. . . large pieces.'

In clauses expressing participant-internal modality, the experiencer argument is both the source of need and the modal subject, undergoing a compulsion (need) to act. Hence, in Hypothesis 1 (next section), we predict that participant-internal

800

modality will correlate with overt expression of the experiencer argument, as this argument has two semantic relations to the proposition and is more relevant for the semantics of the clause. Participant-external modality, on the other hand, is predicted to allow the omission of the experiencer more easily, as the argument is less tightly connected to the core causal dynamics of the clause.

2.5 Hypotheses

The hypotheses guiding the study are the following:

Hypothesis 1

The main factor affecting VARIATION BETWEEN OVERT and NULL EXPRESSION of the experiencer is PARTICIPANT-INTERNAL vs. PARTICIPANT-EXTERNAL MODALITY. The experiencer is more likely to be explicitly expressed in contexts of participant-internal necessity.

Hypothesis 2

The main factor affecting VARIATION BETWEEN ALLATIVE and ADESSIVE MARKING of the experiencer is COMPLEMENT TYPE. Constructions with infinitival complements mainly use adessive case.

3. METHODOLOGY

Our analysis is based on data extracted from the Corpus of Written Estonian and the Corpus of Spoken Estonian. We analysed 605 instances of usages of the 'need'-constructions. The data were coded according to syntactic and semantic factors (predictors) that we considered to be potentially relevant in influencing the choice between adessive, allative or zero marking of the experiencer. We then analysed the predictors using classification and regression trees in combination with random forests to determine the most powerful predictors influencing the expression of the experiencer.

3.1 Data collection and coding

The data for this study were obtained from two corpora, the Corpus of Spoken Estonian (SPO) and the Corpus of Written Estonian (WRI; see Table 1). The spoken language instances of *tarvis* and *vaja* were collected from the Corpus of Spoken Estonian. This mainly includes everyday, face-to-face and telephone conversations, and some institutional conversations; as the corpus consists of short texts with varied speakers, individual speaker preferences will not affect the overall results. Altogether, the spoken data included 248 instances containing either *tarvis* or *vaja*. In the Corpus of Written Estonian, we used the Subcorpus of Fiction from the 1990s. We employed an online search engine (http://www.cl.ut.ee/korp used/kasutajaliides/) to find instances of the *tarvis*/*vaja* 'need'-construction, and obtained a total of 357 instances, containing a nearly equal balance of randomly selected examples with *tarvis* and *vaja*, avoiding the dominance of one author or source.

Corpus	Number of occurrences		Total	%
	tarvis	*vaja*		
Spoken (SPO)	25	223	248	41%
Written (WRI)	179	178	357	59%
Total	204	401	605	100%

Table 1
Overview of data.

Corpus	Zero experiencer		Adessive		Allative		Total	
Spoken (SPO)	152	61.3%	89	35.9%	7	2.8%	248	100%
Written (WRI)	203	56.9%	116	32.5%	38	10.6%	357	100%
Total	355	58.7%	205	33.8%	45	7.4%	605	100%

Table 2
Distribution of zero, adessive and allative experiencer by corpus.

As described in Section 2 above, the modal adverb is the only obligatory element in the construction under investigation. Table 1 shows that in the spoken corpus, *vaja* clearly dominates. In the data drawn from the written corpus, the proportion of *tarvis* and *vaja* clauses is equal; however, *vaja* is reported to be used approximately 3.5 times more frequently than *tarvis* in contemporary written Estonian (Kaalep & Muischnek 2002). Hence, the sample is not representative in terms of the balance in usage of the modal adverbs, *tarvis* and *vaja*, but it was necessary to include a critical amount of clauses with *tarvis* in order to test whether the choice of adverb had an influence on experiencer marking. The data were coded by the authors regarding a number of features, including properties of the clause, properties of the clause constituents and the context of the example. All predictors coded are summarised in Table 3 below and discussed in the next section.

Regarding the two corpora, we expected the experiencer argument to be omitted more often in spoken than in written language because of the generally more elliptical nature of spontaneous oral language (Halliday & Hasan 1976). However, as the distribution in Table 2 shows, the differences in experiencer omission between the data from the spoken and written corpora are not significant ($p > .05$): around 60% omission in both corpora, with spoken data showing only 4.3% more zero experiencers than the written data.

A greater difference between the two corpora emerged, however, in the case-marking of overt experiencers. The examples from the Corpus of Written Estonian

802

attest nearly 8% more experiencers with allative case in 'need'-constructions than the examples from the Corpus of Spoken Estonian, which has remarkably few allative experiencers. This difference is statistically significant ($p < .001$); however, it may also be related to the nature of spontaneous speech, namely the tendency towards apocope, as the phonological distinction between allative and adessive case rests for most words merely in the omission of the final vowel (e.g. *neil* 'they-ADE', *neile* 'they-ALL').[2]

The coding of the experiencer is the phenomenon under investigation, which we further subdivided into two distinct dependent variables: presence of the experiencer argument (*exp_is*, with two levels: overt expression and omission) and case (*exp_case*, with two levels: adessive and allative). The independent predictors hypothesised to influence the expression of the experiencer argument are summarised in Table 3.

3.2 *Coding the independent variables*

In the previous section, we discussed the nature of the corpus and modal adverb predictors. Next we briefly discuss the other independent predictors and provide some examples.

3.2.1 *Main verb (copula verb)*

The verb most commonly used in the 'need'-construction is the copula *olema* 'be'. In addition, two other verbs are also used with a copular function in the construction, albeit rather infrequently: *tulema* 'come' and *minema* 'go' (inflected suppletively as *läinud* in 18a below). The main-verb predictor has four levels according to lexical verb (*olema*, *tulema*, *minema* and ellipsis (0)). Example (18b) shows the construction used with no verb.

(18) (a) Mu detektiivikalduvusi **ei läi-nud** siiski vaja. (WRI)
I.GEN detective.tendencies.PL.PRT not go-APP PRTCL need
'My detective tendencies were not needed after all.'

(b) keegi kinkis jõulu- jõuluks neile triikraua ja siis nemad mõtsid
et **vaja triikimislauda**, sest aint üks köögilaud
that need ironing.table.PRT since only one kitchen.table
väike on (SPO)
small be.3SG
'Someone gave them an iron for Christmas and then they thought they needed an ironing board, because there's only one small kitchen table.'

---

[2] We thank an anonymous *JL* referee for pointing this out in this context. The same referee also noted that the prevalence of adessive in spoken data might be related to different generations of speakers in the spoken and written language corpora. Indeed, this is likely to be the case, and this would provide additional evidence of the ongoing grammaticalisation of the adessive case.

Predictor	Description	Predictor levels
corpus	Corpus containing example	SPO (spoken), WRI(written)
tarvis_vaja	Modal adverb	*tarvis*, *vaja*
mainverb	Copula verb	*olema* 'be', *tulema* 'come', *minema* 'go', 0 (omission)
polarity	Polarity	pos (affirmative), neg (negative)
modality	Type of modality	internal (participant-internal), external (participant-external, including deontic modality)
complement	Complement type	nom (nominal), inf (infinitival), cls (clausal), 0 (no complement)
exp_person	Person of experiencer referent	1sg, 2sg, 3sg, 1pl, 2pl, 3pl, indefinite pronoun, reflexive, 0 (zero person)
RD	Referential distance	1 (preceding clause), 2 (two clauses back), 3 (three or more clauses back), X (referent is not explicitly referred to earlier or not trackable)
LMF	Form of last mention of referent	nom (nominative), ade (adessive), all (allative), other, 0 (no previous explicit mention)

Table 3
Independent variables used in the coding and analysis.

Omission of the copula occurs frequently in Estonian in certain constructions, as is also characteristic of Russian, an important contact language of Estonian. In some contexts, the omitted copula in Estonian may derive from Russian influence; one such context is clauses containing *tarvis*/*vaja* (Kehayov 2009: 129; Lindström et al. 2014). In our data from Standard Estonian, verb ellipsis is rather infrequent in the 'need'-construction overall, but it has a high rate of co-occurrence with ellipsis of the experiencer (84%).

3.2.2 *Polarity*

All clauses were marked for polarity, a binary predictor with two levels, affirmative and negative. Polarity plays an important role in some modal constructions, e.g. in clauses with the verb *tarvitsema* 'to need', where the epistemic meaning appears only in negative clauses (Penjam 2011). Polarity also has some effect on the overt marking of first- and second-person nominative subjects, as the negative verb forms do not show agreement with the subject, and therefore the pronominal subject tends to be overtly expressed more often (Lindström 2010, Sepp 2010). Recall that the copula verb does not inflect for person in the 'need'-constructions, so this may be less relevant here. The copula verb bears the negative marker, and is therefore obligatory (19).

(19) ma üdle-si-n   et   et (.) Üllari-l **ei ole**   vaja
      I   say-PST-1SG that that Üllar-ADE not be.CONNEG need
      midagi   teha.   (SPO)
      anything.PRT do.INF

      'I said that, uh, that Üllar doesn't need to do anything.'

3.2.3 *Modality*

We coded all clauses for participant-external or participant-internal modality (see Section 2.4 above). Coding for this predictor often required referring to the context of the clause in the original text, and was occasionally subject to some ambiguity. We additionally checked the reliability by performing double-blind coding for this variable and discussing any problematic cases. In most cases, the discourse context helped resolve any disagreements. Modality is discussed at greater length in Sections 4 and 5 below.

3.2.4 *Complement*

The 'need'-construction can take a nominal complement marked with partitive case, an infinitival complement, or more occasionally, a finite clausal complement (as shown above, in example (2)). The complement may be omitted, in the event that the object of need or action required is clear from the discourse or general context.

3.2.5 *Experiencer person referent*

Person was marked as 1sg, 2sg, 3sg, 1pl, 2pl, 3pl, interrogative pronoun, reflexive or 0 (referring to generic, 'zero-person' referents; see Kaiser & Vihman 2007, Jokela 2012). Omission of a nominal argument in Estonian can mean either ellipsis of a definite referent or a generalised referent, interpreted according to the construction and context as generic, indefinite, or impersonal. Zero person, by definition, receives zero marking. Both overt and implicit (omitted) arguments

805

were coded for person. In the case of implicit arguments, the intended referent was reconstructed using clause-internal and contextual evidence.

3.2.6 *Referential distance (RD)*

Finally, we examined the referents of the experiencer arguments in their textual context and measured the distance to the previous explicit mention of the same referent. Predictor levels were defined as 1 (preceding clause, or one clause back), 2 (two clauses back), 3 (three or more clauses back), and X (referent is not explicitly referred to earlier or not trackable). The notion 'referential distance' (RD) originally comes from work by Givón (1983), who measured RD in order to analyse the expression of third-person forms in discourse, and has been developed in work on accessibility and salience (e.g. Ariel 1990, Gundel, Hedberg & Zacharski 1993). In Estonian, RD has been shown to play an important role in the expression of first-person forms as well: the shorter the distance, the greater the probability of omitting the first-person pronoun (Lindström et al. 2009). We were interested in whether oblique modal experiencers show the same sensitivity to referential distance as canonical subjects do.

3.2.7 *Form of last mention*

The same considerations led us to also code the case-marking of the last explicit reference in the text to the same argument referent. The coding differentiated between nominative (nom), adessive (ade), allative (all), other and 0 (no previous explicit mention). Here we wished to investigate whether the modal experiencer tends to refer to referents previously encoded as nominative subjects, similarly marked oblique arguments or others. LMF (last mentioned form) indicates whether any priming effects are in operation, i.e. the tendency to repeat the expression of an argument in the same form (see e.g. Travis & Torres Cacoullos 2012). Referents with no earlier mention were coded as 0; these include both impersonal referents (which never encode the subject explicitly) and those whose first mention is in the analysed construction (and hence have no antecedent).

3.3 *Method of analysis*

We constructed two statistical models to find the best fit for our data and uncover the most influential predictors explaining the choice between (i) overt and zero expression of the experiencer and (ii) adessive or allative marking of the overt experiencer. For this, we applied two non-parametric classification methods: recursive partitioning tree models (Hothorn, Hornik & Zeileis 2006) and random forests (Breiman 2001, Strobl et al. 2008). Recursive partitioning in the conditional inference framework (Hothorn et al. 2006) performs recursive binary splitting of the data. The algorithm makes binary splits locally, deciding at each split which variables best classify the data. When there are no more

806

significant predictors for further splitting, the algorithm stops, thus nonsignificant predictors are not included. The advantage of this method is revealing interactions between explanatory variables, as well as providing straightforward visualisations to capture these interactions. Results are represented in tree graphs, such as shown in Figure 1 below (in Section 4.1).

The random forests method (Breiman 2001) provides complementary information to the conditional inference trees. The method constructs a large number of conditional inference trees and, based on these trees, votes for the variables that best classify the data. This allows measurement of the importance of the variables included in the model. Through random permutation of the predictor variable the difference in prediction accuracy before and after each permutation is measured, thus measuring the extent to which the model is improved with the help of the predictor (Strobl et al. 2008). The random forests method works well even in situations with relatively small numbers of observations, large numbers of predictors, and unevenly distributed datapoints (Strobl et al. 2008, Tagliamonte & Baayen 2012), and this method has been successfully implemented in linguistic studies as an alternative to regression models (e.g. Tagliamonte & Baayen 2012, Baayen et al. 2013, Janda 2013).

All computations are made using R (R Development CoreTeam 2013) and we use the *party* package for both conditional inference tree (*ctree*) and random forests analyses (*cforest*). In the next section, we describe the results of the quantitative analysis, followed by discussion and interpretation of results in Section 5.

4. RESULTS

4.1 *Study 1: Explicit expression of the experiencer*

In our first analysis, we were interested in determining which predictors best explain the choice between overt expression or omission of the experiencer argument; thus, the dependent variable in our first model is the presence or absence of the experiencer (*exp_is*). Here, we did not distinguish case-marked alternatives and only looked at the presence or absence of the argument. In the first recursive partitioning tree model (Figure 1), all coded predictors are included in the analysis. However, several levels in the *exp_person* predictor exclude variation by definition: zero person (0) is always absent, while interrogative (int) and reflexive (refl) pronouns are always present. Hence, we excluded exp_person levels: 0, int and refl from the model. The remaining dataset included 493 observations, and the experiencer person predictor (*exp_person*) had six levels (1sg, 2sg, 3sg, 1pl, 2pl and 3pl).

The tree model excludes factors with no effect from the graph. Figure 1 presents only those predictors which were statistically significant in choosing between expression and omission of the experiencer. As Figure 1 shows, the first partition is made by modality: modality is the most significant factor in the overt expression
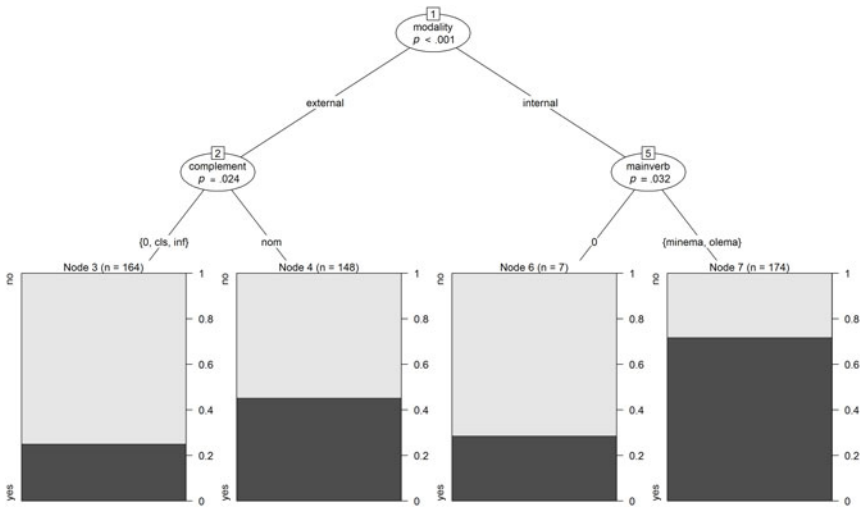
Formula: exp_is ~ tarvis_vaja + mainverb + polarity + complement + exp_person + modality + corpus + LMF + RD

Figure 1
Recursive partitioning tree model for overt expression of the experiencer argument
(excluding from analysis all exp_person levels with no variation).

of the experiencer. Participant-internal modality increases the probability of overt expression of the experiencer. When participant-internal modality is expressed (right branch), the experiencer is explicitly expressed in 70% of clauses (127 out of 181 clauses), while in clauses expressing participant-external modality (left branch), the experiencer is explicitly expressed in 35% (108 out of 312 clauses).

Within the group of participant-external modality (left branch), the type of complement is significant, with the tendency to overtly express the experiencer more often with the nominal complement than with other complement types (compare Node 3 and Node 4). Within the group of participant-internal modality (right branch), the next split is made by the *mainverb* predictor ($p = .032$), partitioning the clauses with no copula into a small group (Node 7, only seven occurrences, with a clear tendency to also omit the experiencer) and clauses with *olema* 'be' and *tulema* 'come' into the second group (Node 8).

Figure 2 presents the variable importance graph obtained from the random forests analysis. The same predictors were included in the model as in the recursive partitioning tree model. In Figure 2, the predictors with the longest bars are the most important, while predictors around zero remain unimportant in explaining the choice between overt and zero experiencer marking. The model shows strong results, with three predictors (*person*, *modality* and *complement*) clearly distinguished as important variables.

Figure 2 shows that person (*exp_person*) has the strongest influence on experiencer omission in this model, followed by modality and complement. All other

808
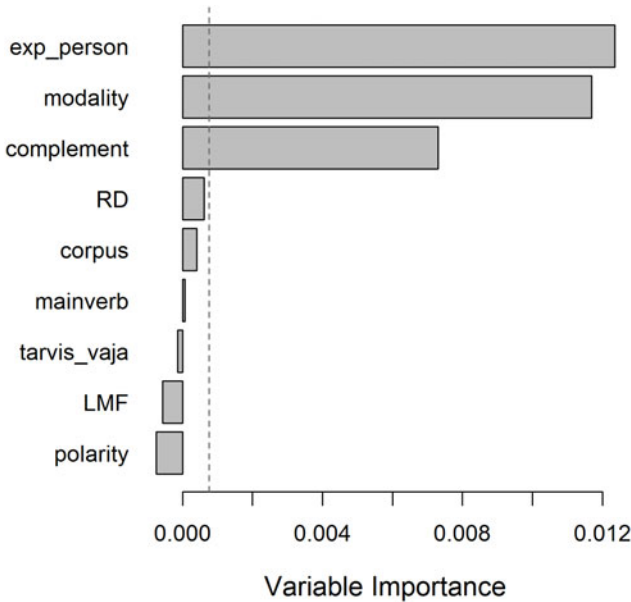
The index of concordance C estimates the goodness of the model (in this case, how well it discriminates between omission and expression of the experiencer argument). In the model, C = .82 and the accuracy of the model is 74%. C ≥ .8 is considered to be a good performance (Harrell 2001: 247).

Figure 2
Variable importance in overt expression of the experiencer argument (random forests analysis). All predictors to the right of the dashed line are important.

predictors remain unimportant in the random forests model. Thus, unlike the recursive partitioning tree model (Figure 1 above), the random forests analysis shows the person of the experiencer as an important predictor, meaning that differences between persons are attested. In Figure 3, we can see that 1sg and 3sg prefer explicit marking of the experiencer, while experiencers tend to be omitted with 3pl. Singular referents in general are less often omitted than plural referents, which may be pragmatically related to shared responsibility or need being less pertinent to express than responsibility or need resting on a single person.

The results regarding expression of the experiencer show that the type of modality was an important predictor in both analyses. This leads us to conclude that our first hypothesis is confirmed: participant-internal modality prefers overtly expressed experiencers.

However, the person referred to is also important; first- and third-person singular experiencer referents are overtly expressed more often, while third-person plural referents often remain unexpressed.

Complement type affected the choice less, but a tendency to express the experiencer overtly with nominal complements was indicated in the analyses.
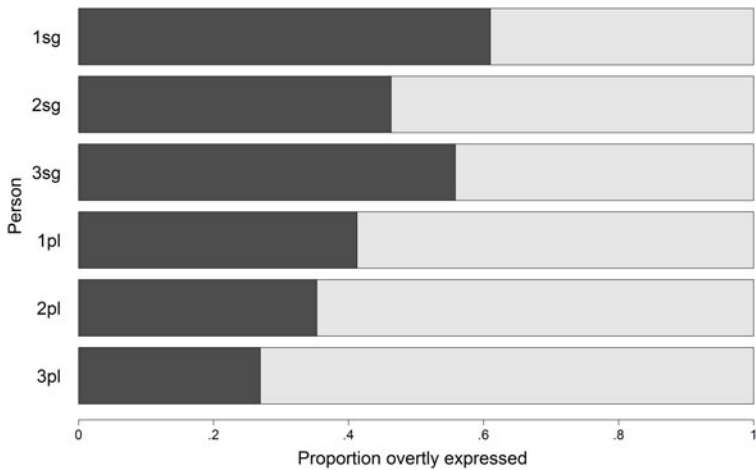
809

Figure 3
Person reference in overt expression of the experiencer argument.

With other types of complement (infinitival, clausal or zero), the experiencer is more often left unexpressed.

In the preceding analysis, the terms 'participant-internal' and 'participant-external' modality were applied to all 'need'-clauses in the data, including premodal constructions, i.e. constructions with nominal complements (as in (2a) above, 'I need a cassette'). This, however, is not unproblematic: the source of need may be less evident in premodal constructions, and more importantly, the internal/external distinction may not be appropriately applied in the case of the premodal constructions, which do not necessarily encode modality at all. To address this issue, we repeated the same analysis with only the 'pure modals', i.e. 'need'-constructions with an infinitival complement. We excluded the *complement* predictor from the analysis; all other independent variables were the same as in the previous analysis. The dataset for this analysis included 217 observations.

The results of the recursive partitioning tree model (Figure 4) are very clear in this analysis: the only significant predictor is *modality*, i.e. participant-internal modality clearly increases the probability of overt experiencer expression in modal 'need'-constructions.

The random forests analysis (Figure 5) indicates that *exp_person* is also an important predictor. Thus, the most important predictors in overt marking of the experiencer are *modality* and *experiencer person*.

810
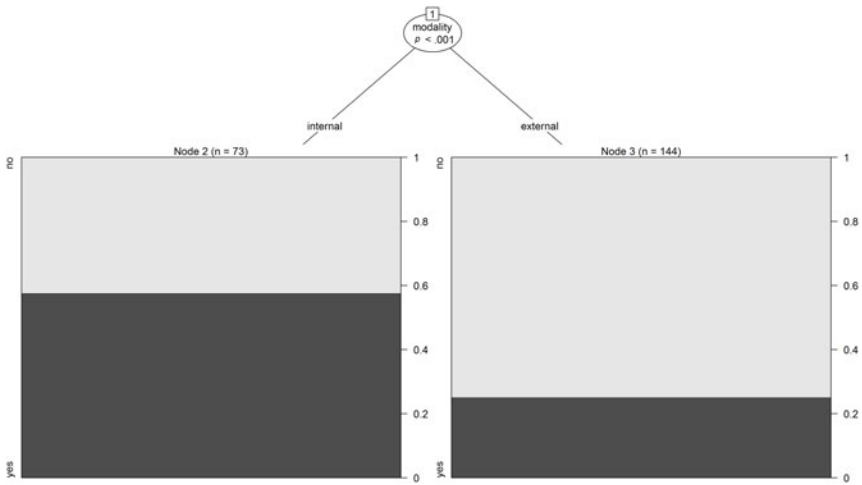
Formula: exp_is ~ tarvis_vaja + mainverb + polarity + exp_person + modality + corpus + LMF + RD

Figure 4
Recursive partitioning tree model for overt expression of the experiencer argument in
modal constructions with infinitival complements (excluding from analysis all exp_person
levels with no variation).
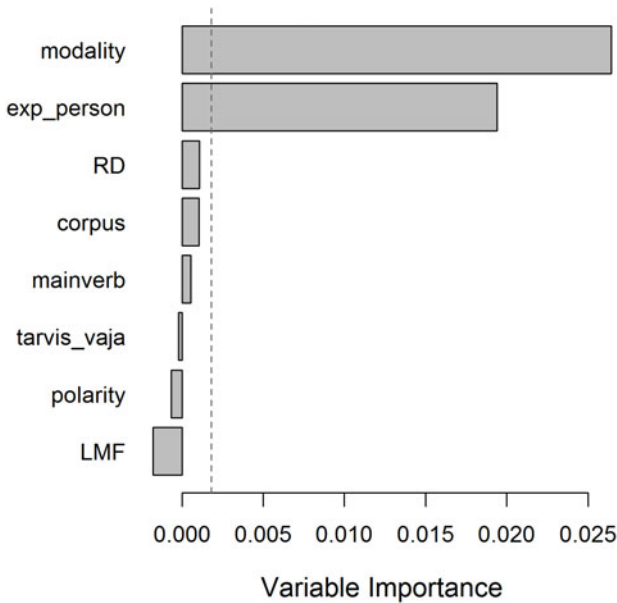
Formula: exp_is ~ tarvis_vaja + mainverb + polarity + exp_person + modality + corpus + LMF + RD; C = .80; accuracy of the model is 76%

Figure 5
Variable importance (overt expression of the experiencer argument in infinitival
'need'-constructions).

811

4.2 *Study 2: Case-marking of the experiencer*

Turning to the case-marking of the experiencer arguments, we included only examples with overt experiencers, marked either with adessive or allative case. In our data, this amounted to 250 examples. As with the analysis for experiencer omission, we used two methods to analyse the data, recursive partitioning tree and random forests. Here, the dependent variable was *exp_case*, with two levels, adessive and allative.
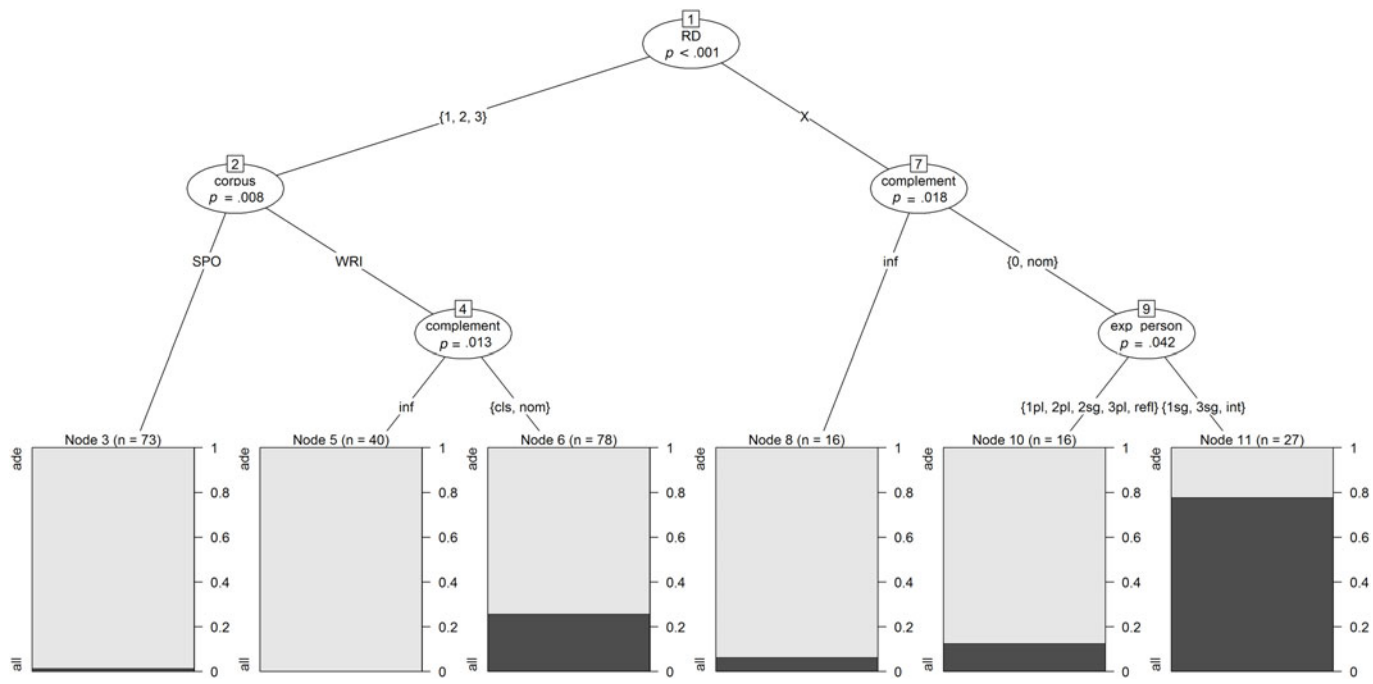
The recursive partitioning tree model in Figure 6 shows that, unlike the previous analysis, the most important predictor in the choice between adessive or allative case-marking is referential distance (*RD*, $p < .001$). Here, a clear distinction is made between experiencer arguments last mentioned within the preceding three clauses and those which are either not previously mentioned or are mentioned earlier. In the right branch (no reference within the preceding three clauses), the allative is used more often than in the left branch. However, as we can see further down, complement type is also an important predictor in both branches: a nominal complement increases the use of allative case, while an infinitival complement almost always only allows adessive case-marking (only one occurrence of allative case in Node 8). Differences also emerged between the spoken and written corpus: the *corpus* predictor makes a significant split in the left branch ($p < .008$), showing a tendency to use more allative experiencers in the written data.

The random forests analysis returns similar results for variable importance in variation in case-marking (Figure 7): person, complement type and referential distance are shown in this model to have almost equal weight. Additionally, form of last mention (*LMF*) and *corpus* also play a role, while *modality* and *mainverb*, which were important in the choice between omission and expression of the experiencer, are not important predictors in this model.

To summarise the results of models accounting for variability in experiencer case-marking, the significant factors affecting the choice between adessive and allative case are:

1. *Person*: Particularly interrogative and indefinite pronouns, which increase the likelihood of allative case-marking, as do third-person singular referents. Referents with other values for person do not significantly affect case-marking.
2. *Complement*: Infinitival complements allow only adessive case-marking, while nominal complements increase the probability of occurrence of allative case-marking.
3. *RD* (referential distance): Particularly level X, signalling new information. Given referents are more likely to be expressed with adessive case, whereas new referents are more likely to be marked with allative case.

Note that *RD* is related to both newness of referents and focus. Allative case is more likely to be used with new and focussed referents, including interrogative

812

Formula: exp_case ~ modality + polarity + complement + tarvis_vaja + mainverb + corpus + RD + LMF + exp_person

Figure 6
Recursive partitioning tree model for case-marking of the experiencer argument.
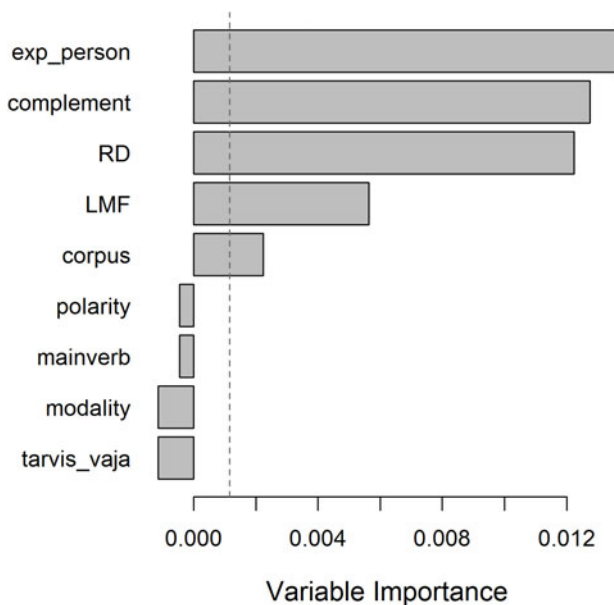
Figure 7
Variable importance (case-marking of the experiencer).

C = .92; accuracy of the model is 86%

pronouns (as exemplified in (20) below), as well as previously unmentioned referents. This is also reflected in the use of allative case for contrastive focus, as in example (8b), above. Allative case is phonologically heavier (one additional syllable), and hence better suited to express focussed constituents.

(20) Kelle-le ol-i   ne-i-d   vaja.
     who-ALL be-PST.3SG they-PL-PRT need
     'Who needed them?'

In addition, the form of last mention of the referent (LMF) may be significant, but different analyses reveal different tendencies, and further study is required. LMF was returned as the fourth predictor of significance in the random forests analysis, but in the recursive partitioning tree model, LMF did not feature as significant. Finally, allative is clearly more common in the written than spoken language.

5. DISCUSSION AND CONCLUSIONS

To investigate variation in expression of the modal experiencer, we examined the data according to two sets of binary choices: first, overt or implicit expression of the experiencer, and second, adessive or allative case-marking of overt arguments.

814

The two most significant predictors for omission of the experiencer – modality and complement – are motivated in different ways. Participant-internal modality affects expression or omission of the experiencer due to semantic reasons. In cases where the need arises from the participant referent him/herself, the experiencer is more intimately connected to the proposition, be it for physiological, psychological, or epistemic reasons. Hence, the expression of that need is more likely to include overt mention of the experiencer. The experiencer referent is both the actor under obligation to act or procure something, as well as the originator of that obligation. In cases of participant-external modality, the referent of the experiencer argument is under obligation to act or procure something, but is not the origin or cause of that need, hence the semantic connection is not as tight, and the experiencer is less likely to be expressed in these clauses. Clauses expressing internal modality are much more likely to have an overt experiencer (over two-thirds of these examples have explicit experiencers, as in (21a)) than clauses with external modality (less than 30%, see (21b)).

(21) (a) ma = t-si-n   et **mu-l ole-ks vaja** nii (.)
           I   say-PST-1SG that I-ADE be-CND need so
           retsepte   **haka-ta saa-ma** mu-l akka-vad
           prescription.PL.PRT start-INF get-SUP I-ADE start-3PL
           rohu-d   otsa saa-ma   (SPO)
           medicine-PL out become-INF

           'I said that I really need to get my prescriptions, my medicine is running out.'

      (b) Kui nende   teema ammendu-s   ja **Mauno-l**
           when they.PL.GEN topic exhaust-PST.3SG and Mauno-ADE
           ol-i   taas **kuhugi   vaja ruta-ta**, oota-s   Ants
           be-PST.3SG again somewhere need rush-INF wait-PST.3SG Ants
           hetke.   (WRI)
           moment.GEN

           'When their topic was exhausted and Mauno needed to rush somewhere again, Ants waited a moment.'

Complement type, on the other hand, may affect the expression or non-expression of the experiencer for syntactic reasons. In the necessive constructions with infinitival complements ('pure modals'), the clause predicates an action (specified in the complement) required of the experiencer argument referent, thereby equating the modal experiencer with an agent or generalised actor. Although the experiencer argument bears oblique case-marking, it is identified with an argument which is canonically mapped to subjects. Constructions with infinitival complements are thus more grammaticalised with abstract relations mapping the null infinitival subject to the experiencer argument of the main

815

clause, giving the modal experiencers properties of subject-like controller arguments. The fact that constructions with infinitival complements are more grammaticalised as modal constructions makes them more likely to include zero arguments, similarly to canonical transitive and intransitive clauses, which may have omitted subjects in Estonian.

Clauses with nominal complements ('premodals'), on the other hand, are semantically and structurally similar to possessor clauses, where both participants – the adessive possessor and the (nominative or partitive) possessee – are typically present in the clause: the omission of the possessor argument has been reported to occur much less often than omission of the experiencer (Metslang 2013). Grammaticalisation in the domain of modality seems to be accompanied by an increase in the omission of the experiencer argument.

Hence, modality and complement, the two main predictors affecting the expression of modal experiencer arguments, have complementary influences. The infinitival complement shows the effects of a grammaticalisation process, over the course of which the experiencer argument has acquired certain subject-like properties. The experiencer has taken more properties of abstract grammatical argument relations, including the ability to control the subject of the infinitival complement. Another, related property is the higher likelihood of omission. Participant-internal modality, on the other hand, emphasises the semantic centrality of the experiencer and thereby increases its subjectivity, making it more likely to be expressed.

Overall, the results regarding the omission of the experiencer argument do not directly suggest subjecthood. If we look more closely at only the examples with overt experiencers, however, we can see important differences. Referential distance does not play a significant role in experiencer argument omission, but it does influence the choice of case-marking. Adessive arguments behave more like subjects: they refer to more salient referents, are preferred in more grammaticalised constructions, and occur in less semantically restricted distribution. The higher the salience, the more likely that the argument will be marked with adessive case. It is also noteworthy that the lower the salience (more distant last mention), the more marked the argument: allative case is more marked both in terms of retaining more of its directional semantics (less bleached of meaning) as well as having more phonological content (an additional syllable, *-le*, compared to the adessive *-l*).

Hence, although topical experiencer arguments do not strongly show the subject-like tendency to be omitted, we instead find that topicality is related to adessive case-marking. This in turn strengthens the claim that adessive case is more grammaticalised in Estonian as a structural, argument-marking case than the more semantic allative case. Adessive experiencers are typically anteceded by nominative arguments referring to the same referent (over 40%), with zero arguments and 'other' following this (at slightly over 20%). Allative experiencers show the opposite pattern, more often introducing a new referent without any antecedent (56%), followed by a quarter of examples with nominative

816

antecedents. Finally, the person of the experiencer referent also plays an important role, but in two distinct ways for the two instances of variation. In experiencer omission, reference to first-person singular has an effect which seems to fit well with internal modality: it is easiest and most typical to speak of one's own inner needs, as speaking of others' internal necessity involves assumptions, beliefs or hearsay (although, in the case of an omnipotent author, the inner state of third-person referents in fiction may be accessible as for first-person referents; this may constitute a difference between our corpora). As for variation in experiencer case-marking, the indefinite and interrogative pronouns play an important role. Here, information structure is likely to affect the picture, as indefinites and interrogatives are often focussed arguments, and for this reason the experiencer argument is likely to be more marked, in addition to always being overt.

The choice of case-marking may also be affected by other factors. Metaphorical directionality (for marking potential recipient-type participants) may increase the likelihood of allative case being used with nominal complements. Constructional homonymy may also play a role. In necessive constructions containing an infinitival complement, it has been pointed out that adessive marking of the experiencer (as in (22a) below) is more typical and clearly preferred, whereas allative marking of the NP invites other interpretations (e.g. recipient or beneficiary, as in (22b); Jaakola 2003: 170–171).

(22) (a) Lapse-l   on   vaja süüa   teha.
          child-ADE be.3SG need eat.INF make.INF
          'The child needs to make food.'

      (b) Lapse-le   on   vaja süüa   teha.
          child-ALL be.3SG need eat.INF make.INF
          'One needs to make food for the child.'

The actor of the event expressed by the infinitive in (22b) is a generic referent, someone other than the 'child' (the allative NP argument); the 'child' referred to by the allative NP is in the role of beneficiary. In coding our data, we did not label noun phrases such as that in (22b) as experiencers, i.e. they were not included in the analysis; however, speakers may make use of adessives to avoid such confusion.

Although the results of our study point to grammaticalisation of the adessive case, it is inaccurate to say that the modal experiencer behaves like a transitive subject (A-argument). Regarding experiencer person, the results of our study do not align with the tendency noted in Estonian to omit first- and second-person more than third-person nominative subjects; rather, the tendency to omit the experiencer is strongest in third-person plural contexts (73%), and lowest in first-person singular (39%). This suggests a role for verb inflection in the tendency to omit nominative first- and second-person referents (leading to the redundancy of pronominal reference), contrasting with the 'need'-constructions, in which the verb does not carry person information.

817

These results also point to the difficulties of assessing the argumenthood of the allative and adessive arguments. In terms of a cross-linguistic typology, the distinction between arguments and adjuncts is difficult to make in a general, principled way (see e.g. Creissels 2014, Haspelmath 2014). The arguments we examine in this paper are not subject-like in terms of their coding properties. Nevertheless, the predicate semantics and discourse pragmatics strongly suggests argumenthood: they are lexically encoded by the predicate, required to bear certain properties, and their absence is interpreted as omission or generalised reference: i.e. they are semantically obligatory, even when syntactically absent. Successful interpretation of the predicate requires interpretation of oblique arguments. Oblique adjuncts, on the other hand, are not obligatory, can be freely omitted, and hence when they are absent, they do not give rise to any specific, obligatory interpretation. Neither their presence nor their form is determined by predicate semantics. In the case of oblique arguments which are more subject-like, they are also freely omitted in Estonian, but this leads to consequences in interpretation and resolution. The omission of more subject-like oblique arguments is conditioned by discourse pragmatics, but leads to a paradoxical difficulty for coding and analysis. Their more frequent omission indicates not optionality, but a stronger semantic presence.

It is worth noting that Estonian shares some of the patterns of experiencer argument expression with the closely related language Finnish, yet with important differences. In Finnish, omission of the experiencer argument is very common in necessary constructions, but overt experiencers are marked mainly with genitive case, unlike Estonian (Laitinen 1992, Laitinen & Vilkuna 1993). The elliptical NP refers to people familiar from the context, either the discourse participants or the narrative protagonist (see Laitinen 1992: 109–110). Experiencer omission in necessary constructions referring to discourse participants has been noted in the Finnish academic grammar (ISK: 1291). Interestingly, however, Laitinen finds that, at least in dialect interviews, omission is used mainly for generalising, non-referential arguments ('zero person', Laitinen 1992: 110). According to our results, the generalising zero person is used similarly but less frequently in Estonian. Investigations of dialect data have shown more experiencer omission than in the standard language analysed here, which may also reflect differences between genres (Lindström et al. 2014).

Finally, the very fact of such high rates of omission is worth emphasising, as oblique NPs have often been likened formally to adjuncts, and therefore their realisation as zero is unlikely, or in other words, would be interpreted as their absence (Siewierska 2003). The finding that certain oblique arguments can be omitted so frequently, while being given a referential interpretation, has consequences: for speakers making choices regarding expression of arguments, theorists making claims about tendencies and typology, and, last but not least, for the process of data coding and analysis, and the care and consideration which must be taken not to overlook invisible elements.

818

ABBREVIATIONS

0	zero person
1, 2, 3	first, second, third person
ABE	abessive
ADE	adessive
ALL	allative
APP	active past participle
CLI	clitic (=)
CND	conditional
COM	comitative
CONNEG	connegative
ELA	elative
GEN	genitive
ILL	illative
INF	infinitive
int	interrogative pronoun
IPS	impersonal
LMF	last mentioned form
NEG	negation
NOM	nominative
PL	plural
PPP	passive past participle
PRS	present tense
PRT	partitive
PRTCL	particle
PST	simple past
RD	referential distance
REFL	reflexive pronoun
SG	singular
SPO	spoken Estonian corpus
SUP	supine
TER	terminative
TRL	translative
WRI	written Estonian corpus

REFERENCES

Ariel, Mira. 1990. *Accessing NP antecedents*. London: Routledge & Croom Helm.

Baayen, R. Harald, Anna Endresen, Laura A. Janda, Anastasia Makarova & Tore Nesset. 2013. Making choices in Russian: Pros and cons of statistical methods for rival forms. *Russian Linguistics* 37.3, 253–291.

Bickel, Balthasar. 2004. The syntax of experiencers in the Himalayas. In Peri Bhaskararao & Karumuri Venkata Subbarao (eds.), *Non-nominative subjects* (Typological Studies in Language 61), vol. 2, 77–111. Amsterdam: John Benjamins.

Blansitt, Edward L. 1988. Datives and allatives. In Michael Hammond, Edith A. Moravcsik & Jessica Wirth (eds.), *Studies in syntactic typology* (Typological Studies in Language 17), 173–191. Amsterdam: John Benjamins.

Bolinger, Dwight. 1980. Wanna and the gradience of auxiliaries. In Gunter Brettschneider & Christian Lehmann (eds.), *Wege zur Universalien Forschung*, 292–299. Tübingen: Gunter Narr Verlag.

Bossong, Georg. 1998. Le marquage de l'expérient dans les langues de l'Europe. In Jack Feuillet (ed.), *Actance et valence dans les langues de l'Europe*, 259–294. Berlin: Mouton de Gruyter.

Breiman, Leo. 2001. Random forests. *Machine Learning* 45.1, 5–32.

Creissels, Denis. 2009. Spatial cases. In Andrej Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 609–625. Oxford: Oxford University Press.

Creissels, Denis. 2014. Cross-linguistic variation in the treatment of beneficiaries and the argument vs. adjunct distinction. *Linguistic Discovery* 12.2, 41–55.

Croft, William. 1993. Case marking and the semantics of mental verbs. In James Pustejovsky (ed.), *Semantics and the lexicon*, 55–72. Dordrecht: Kluwer.

de Hoop, Helen & Peter de Swart (eds.). 2009. *Differential subject marking*. Dordrecht: Springer.

Duvallon, Outi & Antoine Chalvin. 2004. La réalisation zéro du pronom sujet de première et de deuxième personne du singulier en finnois et en estonien parlés. *Linguistica Uralica* XL.4, 270–286.

EKG II = Mati Erelt, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael & Silvi Vare. 1993. *Eesti keele grammatika II. Süntaks* [Grammar of Estonian II: Syntax]. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.

Erelt, Mati. 2013. *Eesti keele lauseõpetus. Sissejuhatus. Öeldis* [Estonian sentence structure. Introduction: The predicate]. Tartu: Preprints of the Department of Estonian of the University of Tartu 4.

Erelt, Mati & Helle Metslang. 2006. Estonian clause patterns: From Finno-Ugric to standard average European. *Linguistica Uralica* XLII.4, 254–266.

Erelt, Mati & Metslang Helle. 2008. Kogeja vormistamine eesti keeles: nihkeid SAE perifeerias [Expression of the experiencer in Estonian: Shifts in the periphery of SAE]. *Emakeele Seltsi aastaraamat* [Yearbook of the Estonian Mother Tongue Society] 53, 9–22.

Givón, T. 1983. Introduction. In T. Givón (ed.), *Topic continuity in discourse: A quantitative cross-language study*, 5–41. Amsterdam & Philadelphia, PA: John Benjamins.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69.2, 274–307.

Halliday, M. A. K. & Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.

Hansen, Björn. 2014. The syntax of modal polyfunctionality revisited: Evidence from the languages of Europe. In Elisabeth Leiss & Werner Abraham (eds.), *Modes of modality: Modality, typology, and universal grammar* (Studies in Language Companion Series 149), 89–126. Amsterdam & Philadelphia, PA: John Benjamins.

Harrell, Frank E. Jr. 2001. *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. New York: Springer.

Haspelmath, Martin. 2001. Non-canonical marking of core arguments in European languages. In Aikhenvald et al. (eds.), 53–83.

Haspelmath, Martin. 2014. Arguments and adjuncts as language-particular syntactic categories and as comparative concepts. *Linguistic Discovery* 12.2, 3–11.

Heine, Bernd. 1993. *Auxiliaries: Cognitive forces and grammaticalization*. New York & Oxford: Oxford University Press.

Hint, Helen. 2015. Third-person pronoun forms in Estonian in the light of Centering Theory. *Journal of Estonian and Finno-Ugric Linguistics* 6.2, 105–135.

Holvoet, Axel. 2007. *Mood and modality in Baltic*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.

Hothorn, Torsten, Kurt Hornik & Achim Zeileis. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15.3, 651–674.

ISK = Auli Hakulinen, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen & Irja Alho. 2004. *Iso suomen kielioppi* [Comprehensive grammar of Finnish]. Helsinki: Suomalaisen Kirjallisuuden Seura.

Jaakola, Minna. 2003. Kokijarakenteen kontrastointia [Contrasting experiencer constructions]. In Pirkko Muikku-Werner & Hannu Remes (eds.), *Viro ja suomi: kohdekielet kontrastissa. Lähivertailuja* [Estonian and Finnish: Contrasting target languages. Close comparisons] 13, 167–177.

Janda, Laura A. 2013. Quantitative methods in Cognitive Linguistics: An introduction. In Lura A. Janda (ed.), *Cognitive linguistics: The quantitative turn. The essential reader*, 1–32. Berlin & Boston, MA: De Gruyter Mouton.

820

Jokela, Hanna. 2012. *Nollapersoonalause suomessa ja virossa. Tutkimus kirjoitetun kielen ainestosta* [Zero-person clauses in Finnish and Estonian: A study on written language] (Annales universitatis turkuensis, ser. C tom. 334). Turku: University of Turku.

Kaalep, Heiki-Jaan & Kadri Muischnek. 2002. *Eesti kirjakeele sagedussõnastik* [Frequency dictionary of Standard Estonian]. Tartu.

Kaiser, Elsi & Virve-Anneli Vihman. 2007. Invisible arguments: Effects of demotion in Estonian and Finnish. In Benjamin Lyngfelt & Torgrim Solstad (eds.), *Demoting the agent: Passive, middle and other voice phenomena*, 111–141. Amsterdam: John Benjamins.

Kehayov, Petar. 2009. *Olema*-verbi ellipsist eesti kirjakeeles [Ellipsis of the copula in Standard Estonian]. *Emakeele Seltsi aastaraamat* 54, 107–152.

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 & New York: Mouton de Gruyter.

Klavan, Jane, Kaisa Kesküla & Laura Ojava. 2011. Synonymy in grammar: The Estonian adessive case and the adposition *peal* 'on'. In Seppo Kittilä, Katja Västi & Jussi Ylikoski (eds.), *Studies on case, animacy and semantic roles*, 113–134. Amsterdam: John Benjamins.

Koenig, Jean-Pierre, Gail Mauner & Breton Bienvenue. 2003. Arguments for adjuncts. *Cognition* 89, 67–103.

Kuteva, Tania. 2001. *Auxiliation: An enquiry into the nature of grammaticalization*. Oxford: University Press.

Laitinen, Lea. 1992. *Välttämättömyys ja persoona. Suomen murteiden nesessiivisten rakenteiden semantiikkaa ja kielioppia* [Necessity and person: The semantics and grammar of necessive structures in Finnish dialects]. Helsinki: Suomalaisen Kirjallisuuden Seura.

Laitinen, Lea & Maria Vilkuna. 1993. Case-marking in necessive constructions and split intransitivity. In Anders Holmberg & Urpo Nikanne (eds.), *Case and other functional categories in Finnish syntax* (Studies in Generative Grammar 39), 23–48. Berlin & New York: Mouton de Gruyter.

Lestrade, Sander. 2010. *The space of case*. Ph.D. dissertation, Radboud Universiteit Nijmegen.

Lindström, Liina. 2010. Kõnelejale ja kuulajale viitamise vältimise strateegiaid eesti keeles [Strategies of avoidance of reference to the speaker and hearer in Estonian]. *Emakeele Seltsi aastaraamat* 55, 88–118.

Lindström, Liina. 2013. Between Finnic and Indo-European: Variation and change in the Estonian experiencer-object construction. In Ilja A. Seržant & Leonid Kulikov (eds.), *The diachronic typology of non-canonical subjects*, 141–164. Amsterdam & Philadelphia, PA: John Benjamins.

Lindström, Liina. 2015. Subjecthood of the agent argument in Estonian passive constructions. In Marja-Liisa Helasvuo & Tuomas Huumo (eds.), *Subjects in constructions – canonical and non-canonical*, 141–173. Amsterdam & Philadelphia, PA: John Benjamins.

Lindström, Liina, Mervi Kalmus, Anneliis Klaus, Liisi Bakhoff & Karl Pajusalu. 2009. Ainsuse 1. isikule viitamine eesti murretes [First-person singular reference in Estonian dialects]. *Emakeele Seltsi aastaraamat* 54, 159–185.

Lindström, Liina & Ilona Tragel. 2010. The possessive perfect construction in Estonian. *Folia Linguistica* 44.2, 371–399.

Lindström, Liina, Kristel Uiboaed & Virve-Anneli Vihman. 2014. Varieerumine *tarvis/vaja*-konstruktsioonides keelekontaktide valguses [Variation in *tarvis/vaja* constructions in the light of language contact]. *Keel ja Kirjandus* 8–9, 609–630.

Metslang, Helena. 2013. Coding and behavior of Estonian subjects. *Journal of Estonian and Finno-Ugric Linguistics (ESUKA – JEFUL)* 4.2, 217–293.

Narrog, Heiko. 2010. Voice and non-canonical case marking in the expression of event-oriented modality. *Linguistic Typology* 14, 71–126.

Næss, Åshild. 2007. *Prototypical transitivity* (Typological Studies in Language 72). Amsterdam: John Benjamins.

Needham, Stephanie & Ida Toivonen. 2011. Derived arguments. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG11 Conference*. Stanford, CA: CSLI Publications, http://csli-publications.stanford.edu/.

Onishi, Masayuki. 2001. Non-canonically marked subjects and objects: Parameters and properties. In Aikhenvald et al. (eds.), 1–51.

Pajusalu, Erna. 1958. Adessiivi funktsioonid eesti murretes ja lähemates sugulaskeeltes [Functions of the adessive case in Estonian dialects and closely related languages]. *Keel ja Kirjandus* 4/5, 246–258.

Penjam, Pille. 2006. Tulema-verbi grammatilised funktsioonid eesti kirjakeeles [Grammatical functions of the verb *tulema* 'come' in Standard Estonian]. *Keel ja Kirjandus* 1, 33–41.

Penjam, Pille. 2011. Eesti kirjakeele subjektilised ja adessiivadverbiaaliga *tarvitsema*-konstruktsioonid [*Tarvitsema* 'need'-constructions with subjects and adessive adverbials in Standard Estonian]. *Keel ja Kirjandus* 7, 505–525.

R Development CoreTeam. 2013. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Accessible at http://www.R-project.org/.

Ross, Kristiina. 1997. Kohakäänded Georg Mülleri ja Heinrich Stahli eesti keeles [Local cases in the Estonian of Georg Müller and Heinrich Stahl]. In Mati Erelt, Meeli Sedrik & Ellen Uuspõld (eds.), *Pühendusteos Huno Rätsepale* [Festschrift for Huno Rätsep], *28.12.1997*. Tartu Ülikooli eesti keele õppetooli toimetised 7, 184–201.

Sepp, Pille. 2010. Pronoomeni kasutus MSN-vestlustes [The use of pronouns in MSN chats]. BA thesis, Department of Estonian, University of Tartu.

Siewierska, Anna. 2003. Reduced pronominals and argument prominence. In Miriam Butt & Tracy Holloway King (eds.), *Nominals: Inside and out*, 119–150. Stanford, CA: CSLI Publications.

Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin & Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9.1, 307.

Tagliamonte, Sali A. & 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.

Travis, Catherine E. & Rena Torres Cacoullos. 2012. What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation. *Cognitive Linguistics* 23.4, 711–748.

Tutunjian, Damon & Julie E. Boland. 2008. Do we need a distinction between arguments and adjuncts? Evidence from psycholinguistic studies of comprehension. *Language and Linguistics Compass* 2, 631–646.

van der Auwera, Johan & Vladimir A. Plungian. 1998. Modality's semantic map. *Linguistic Typology* 1.2, 79–124.

van der Auwera, Johan, Petar Kehayov & Alice Vittrant. 2009. Acquisitive modals. In Lotte Hogeweg, Helen de Hoop & Andrej Malchukov (eds.), *Cross-linguistic semantics of tense, aspect and modality*, 271–302. Amsterdam & Philadelphia, PA: John Benjamins.

Zinken, Jörg & Eva Ogiermann. 2011. How to propose an action as objectively necessary: The case of Polish *trzeba x* ('one needs to x'). *Research on Language and Social Interaction* 44(3), 263–287.

Authors' addresses: *Institute for Estonian and General Linguistics, University of Tartu, Jakobi 2, 51014 Tartu, Estonia*
liina.lindstrom@ut.ee
virve.vihman@ut.ee