OBJECT CASE VARIATION OF THE PRONOUN *MIS* 'WHAT' IN SPONTANEOUS SPOKEN ESTONIAN AND ESTONIAN DIALECTS

Hanna Pook
University of Tartu, EE
hanna.pook@ut.ee

Abstract. The Estonian language makes a systematic distinction between total and partial objects on the basis of semantic and syntactic features: total objects occur in nominative or genitive, partial objects in partitive. However, in the case of the interrogative-relative pronoun *mis* 'what', the partitive *mida* in the expected partial object position can be replaced with the nominative *mis*. The aim of this paper is to determine which variables significantly affect this object case variation, how the variation differs between contemporary speech and archaic dialects and what might have possibly motivated the development of this variation. This study is based on the data in the Phonetic Corpus of Estonian Spontaneous Speech and the Corpus of Estonian Dialects. The results show that the variation is most affected by verb type, clause type, length of the following word and dialect. It is concluded that there might be multiple motivations behind this variation, mainly language contact (or a lack of it in certain areas), high usage frequency of the pronoun *mis* and the effect of the standardisation of language.

1. Introduction

1.1. General background

Estonian and other Finnic languages exhibit a phenomenon called differential object marking (DOM) which entails making a systematic distinction between total and partial objects on the basis of semantic and syntactic features (Aissen 2003, Witzlack-Makarevich & Seržant 2018). In Estonian, the choice between using a total or a partial object depends mostly on polarity, aspect and the referent's boundedness. In

order to use a total object, all three of the following conditions must be met: the situation described in a sentence is perfective, the referent is quantitatively bounded and the polarity of the sentence is affirmative. Otherwise, a partial object is used (Erelt et al. 1993: 51–52, Metslang 2017: 264–267, see also Kont 1963, Ogren 2015).

However, there are a few exceptions to object form variation (see Metslang 2017: 272–273) and the interrogative-relative pronoun *mis* 'what' is one of them: in a position where a partial object is expected to occur[1], the partitive *mida* can be replaced with the nominative *mis* (1a–b). This variation seems to occur in different clause types (questions, relative clauses, indirect questions, etc.) as well as in written, spoken, dialectal, and even old literary Estonian. In these clauses, *mis* can be interpreted as a grammatical object, similarly to *mida* (Pajusalu 2006: 330).

(1) a. ***Mida*** ~   ***mis***   *sa*   *teed?*
         what:PRT ~   what:NOM   you   do:2SG
         'What are you doing?'

     b. *Ma*   *ei*   *tea,*   ***mida***   ~ ***mis***   *teha.*
         I   not   know:CNG   what:PRT   ~ what:NOM   do:INF
         'I don't know what to do.'

Similar variation exists in Livonian, where the nominative *mis* has taken on the functions of the partitive *midā*, so much so that the partitive is rarely, if ever, used. In fact, in Salaca Livonian, *mis* can alternate with most cases, not only the partitive (Viitso & Ernštreits 2012, Winkler & Pajusalu 2018: 96).

When considering this case variation from a more general perspective, examples in the dialectal descriptions of the Northeastern, Coastal and Eastern dialects as well as the Risti subdialect of the Western dialect show that a nominative total object is often used instead of the expected partitive partial object, sometimes even in sentences with negative

---

1   Henceforth, the position where one would expect a partial object to occur (based on previous knowledge of Estonian grammar) is simply called *the partial object position* for brevity's sake. However, it is incorrect to claim that this slot only allows for a partitive partial object, when in some instances (like for the pronoun *mis*) there is no partitive in that position at all.

polarity (Kont 1963: 115, Juhkam 1983: 122–123, Alvre 1986: 7, Must 1987: 287). The partitive case can also be substituted by nominative or accusative for all nouns in partial object position in Veps, Ludic and sporadically Karelian; in those languages the use of a total object case in this position has attained dominance compared to using the partitive (Kont 1963: 46–47, 107–109).

Opposite examples where a partial object is used instead of an expected total object are also evident in Finnic languages (including Estonian). In the case of the personal pronouns *mina* 'I', *sina* 'you', *meie* 'we' and *teie* 'you (pl)' and reflexive pronouns *(ise)enese ~ (ise)enda* 'one's own' in Standard Estonian, the partitive case is almost always used, although according to the general DOM rules, a total object is expected in this position (2a–b) (Metslang 2017: 272–273). This exception in DOM for personal pronouns is characteristic to most Finnic languages, but is more frequent in Southern Finnic (Kont 1963: 103).

(2)   a. *Sõbrad*   *jätsid*   **meid**   *üksinda.*
        friend:PL   leave:PST:3PL   we:PRT   alone
        'Friends left us alone.'

     b. *Ta*   *võttis*   **ennast**   *kokku.*
        s/he   take:PST:3SG   him/herself:PRT   together
        'He pulled him/herself together.'

In Finnic languages that have been under strong Russian influence (mainly Votic, Karelian and Veps), partitive can be used instead of the expected total object cases when the object represents an animate being, but this phenomenon is not systematic and is limited to sporadically occurring examples (Kont 1963: 101–102).

In studies on standard spoken Estonian that have dealt with *mis*-initial questions, the nominative *mis* and the partitive *mida* in partial object position have been treated as somewhat analogous and the researchers have not really focused on the difference in the use of the two case forms (see e.g. Pajusalu 2006, Rumm 2019). However, Laanesoo (2014: 121–122) found that prosodically there is a slight difference between questions starting with *mis* and those starting with *mida*

when the question was meant to be a command and not a way to ask for information.[2]

It is interesting to note here that *kes* 'who' in partial object position is always used in the partitive case without variation. This kind of contrast between *mis* and *kes* is, however, not at all unusual, since there are many Indo-European languages in which the interrogative and/or relative *who* has (partly) retained its declension, while *what* has not, e.g. the English *who*, *whom*, *whose* vs. *what/that* and the German *wer* 'who', *wen* 'who(m)', *wessen* 'whose' vs. *was* 'what' (Berry 2018: 122, Durrell et al. 2013: 43), or does not differentiate between nominative and accusative, e.g. the Russian *кто* 'who', *кого* 'whom' vs. *что* 'what' (Timberlake 2004: 117).

It is difficult to say for certain why this exception in object case exists for the pronoun *mis*. In this paper I propose three possible explanations for the emergence and development of the *mis* ~ *mida* alternation: language contacts, high usage frequency and language standardisation. A detailed discussion about the possible influences of those three factors in the light of the paper's statistical analysis is presented in Section 4.

1.2. Aim of the present study

In this study I take a closer look at the alternation between *mis* and *mida* in partial object position in order to explain the essence of and the motivation behind this variation. Typically, the case of the object also changes the meaning of the sentence (Metslang 2017: 259). Therefore, if the nominative *mis* and the partitive *mida* in partial object position were completely interchangeable, the use of this pronoun as an object would significantly differ from the use of other nouns. The main aim of this study is to show, however, that there are certain variables that do

---

2   In many languages, questions, in addition to having the primary function of asking for information, can have many other functions, like voicing a command, a reproach, a complaint, an accusation, a protest, etc. (de Ruiter 2012, Ilie 2015). For example, the question *What are you looking at?* could be interpreted as a straight-forward request to know what the listener is looking at, but also as a directive for the listener to stop looking at the speaker, depending on the context and intonation of said question. The same principle is true in Estonian, where questions (including *mis*-initial questions) can function in a variety of ways and oftentimes fill more than one pragmatic function in a conversation (see Hennoste, Rääbis & Laanesoo 2017, Laanesoo 2018, Rumm 2019).

explain and affect the variation between the two cases in partial object position. The variables examined in the study include both those which are considered relevant in the standard Estonian grammatical descriptions as well as others that I have observed to have a possible influence on this variation.

The data used for this study comes from the Phonetic Corpus of Estonian Spontaneous Speech[3] and the Corpus of Estonian Dialects[4]. I have included both contemporary spontaneous speech and traditional colloquial dialects in my study first and foremost in order to compare the two sets of language data in their use of the pronoun *mis* in partial object position, but also in order to discern the possible motivations behind this variation, as these two sets of language data exhibit considerable differences for example in terms of language contact and language standardisation.

The research questions are the following:

1. Which of the considered variables (clause type, polarity, tense, use of impersonal voice, length of the following word, verb type, dialect, speaker) correlate with the speakers' choice between *mis* and *mida* in partial object position in both spontaneous speech and dialects?
2. How do dialects and spontaneous speech differ in terms of the variation of *mis* and *mida* in partial object position? What might cause the possible differences?
3. Can this variation be explained by contacts with one or multiple languages, high usage frequency of the pronoun, standardisation of language, or something else entirely?

This paper is structured as follows: In Sections 2.1 and 2.2, I describe the corpora, the datasets and data coding. Section 2.3 gives an overview of the methods applied to the data. Section 3 presents the results of the analysis: Section 3.1 includes the results for the spontaneous speech data, Section 3.2 for the dialect data; Section 3.3 shows the analysis of a combined dataset, compares the results of the two corpora and elaborates on the significant variables. In Section 4, I discuss the possible motivations behind this variation. Section 5 contains the conclusions.

---

3   https://doi.org/10.15155/1-00-0000-0000-0000-001A3L (Accessed June 20, 2019.)

4   https://doi.org/10.15155/1-00-0000-0000-0000-00076L (Accessed January 25, 2019.)

2. Data and methods

2.1. Corpora

The data for this paper are derived from two corpora: the Phonetic Corpus of Estonian Spontaneous Speech (PCESS) and the Corpus of Estonian Dialects (CED). Both corpora feature spoken language and were specifically chosen in order to study natural and unedited speech.

PCESS consists of spoken texts that have been recorded since 2006 and it is continuously being updated with new recordings. The speakers represent a variety of ages, genders and social and regional backgrounds. The corpus includes both spontaneous dialogues between speakers, who typically already know each other well, and semi-spontaneous formal monologues in the form of presentations, lectures, etc.

CED consists of spoken dialect texts recorded mostly during the 1960s and 1970s in all Estonian dialect areas. The recorded speakers are usually local, older people (born mostly in 1870–1890), are typically not very highly educated and have not moved around much during their lives. The conversations (which often include long passages of monologues by the speaker) feature topics such as the speaker's personal life and lifestyle, past events and working methods. In the corpus, the dialects have been divided into ten traditional dialect areas: Mid, Western, Insular, Eastern, Coastal, Northeastern, Tartu, Võru, Mulgi and Seto dialects. These dialect areas are depicted on Figure 1.

The total number of speakers, tokens and the lemma *mis* in both corpora is presented in Table 1. It should be noted that while both of these corpora contain spoken language, the languages in them are in essence quite different. The speakers in CED speak a vernacular language that is nowadays archaic, while the speakers in PCESS speak contemporary Estonian. Linguistically, the dialects can be quite different from each other (although, by the time of the recordings, the levelling influence of the standard language had definitely changed them from their original state into a more homogeneous language), while the spontaneous speech texts feature language that is more or less uniform across all speakers.

Figure 1. Estonian dialects in the CED.

Table 1. The number of speakers, morphologically annotated tokens and the frequency of the lemma *mis* in PCESS and CED.

CORPUS	SPEAKERS	TOTAL TOKENS	LEMMA *MIS*
PCESS	121	685,750	4,363
CED	277	1,229,084	7,558
Σ	398	1,914,834	11,921

2.2. Data and coding

For this study, I collected all the uses of the pronoun *mis* from the morphologically annotated corpus texts and compiled the pronouns into datasets with a total of 4,363 observations from PCESS and 7,558 observations from CED. Every observation is accompanied by the preceding and following context (up to 20 words) and information about the speaker.

In this study I focus on the variation of *mis* and *mida* that takes place in the object position where, according to standard Estonian DOM rules, only a partial object is expected. Therefore, I have included in

the analysis only those sentences in which the pronoun *mis* is either in nominative or partitive and in a position where one would expect a partitive object to occur. This selection is based on contemporary Estonian grammar descriptions and my own native-speaker intuition.

After manually eliminating all the sentences in which *mis* was not in partial object position, the final datasets consisted of 1,033 observations from PCESS and 1,481 observations from CED. The distribution of the nominative *mis* and the partitive *mida* in both datasets can be seen in Table 2.

Table 2. Nominative and partitive distribution of the pronoun *mis* in partial object position.

CORPUS	*MIS*	*MIDA*	Σ
PCESS	396 (38.3%)	637 (61.7%)	1,033
CED	1,360 (91.8%)	121 (8.2%)	1,481

In dialects, the nominative *mis* is dominant in partial object position, while in spontaneous Estonian, the distribution of the two cases is much more equal. While for certain constructions it has been argued that sentences starting with *mida* can have an archaic style and are uncommon in contemporary language (e.g. elative *wh*-constructions in Pajusalu 2006: 328–329), it is clear from this distribution that this cannot be true for all possible constructions.

For each instance of *mis* in partial object position, the following variables were either coded or retrieved from the corpus annotations. A summary of all the variables, their possible values and abbreviations used in the subsequent graphs are also presented in Table 3.

Case-marking of *mis*. This is the dependent variable of the study and marks whether the pronoun *mis* in partial object position is in nominative (3a) or partitive (3b). For the dialect dataset, the nominative value also includes the variants *miä ~ mia ~ meä*, which represent the old nominative in the southern dialects (Alvre 1987: 30). Although these are generally considered to be separate from *mis*, the use of them in the data is marginal (only eight occurrences), which is why they are combined with the general nominative used in Standard Estonian and the northern dialects.

(3) a. CED (Tartu)[5]

mis	*te*	*siss*	*nüüd*	*viil*	*tahattõ*
what:NOM	you:PL	then	now	more	want:2PL

'What more do you want now?'

  b. PCESS

artiklite	*kvaliteet*	*ei*	*ole*	*päris*	*see*
article:PL:GEN	quality	not	be:CNG	quite	it

mida	*me*	*tahaksime*
what:PRT	we	want:COND:1PL

'The quality of the articles is not quite what we would want it to be.'

Clause type. This marks whether the pronoun *mis* appears in a relative clause (4a), a question (4b), a rhetorical question (4c), an indirect question (4d) or a relative clause without a main clause[6] (4e). These clause types differ from each other concerning the focus of the respective sentences: in questions and indirect questions, the question word *mis* itself is almost always the focus, but in relative clauses, *mis* takes the role of the topic (Lindström 2006: 881–883). This could possibly affect the choice of the case of *mis*. Clause type also proved to be highly significant in a study about variation in the use of the pronoun *kes* 'who' in Estonian dialects (Pook 2019) (although the focus of that study was not object case, but the referent's animacy, since, as mentioned, *kes* does not have case variation in this position).

(4) a. CED (Mulgi)

ma	*kõnele*	*sedä*	***miss***	*miu*	*esä*
I	speak:1SG	this:PRT	what:NOM	me:GEN	father

omm	*miul*	*kõnelnu*
be:3SG	me:ADE	speak:PST:PTCP

'I speak of what my father has told me.'

---

5 This and all the following examples are derived from the two corpora used in this study: the Phonetic Corpus of Estonian Spontaneous Speech and the Corpus of Estonian Dialects, which are marked before each example as PCESS and CED, respectively. For the dialect sentences, the dialect also precedes the example in parentheses. A brief overview of these corpora has been presented in Section 2.1.

6 This is a construction characteristic of spoken language, which either has the function of introducing a new topic, is an answer to the interviewer's question or lacks a main clause simply because the speaker's train of thought changed mid-sentence.

b. PCESS

aga	***mis***	*sa*	*näiteks*	*seal*	*PÖFFil*
but	what:NOM	you	for_example	there	PÖFF:ADE

vaatasid
watch:PST:2SG

'What did you, for example, watch at the PÖFF film festival?'

c. CED (Western)

mis	*seal*	*muud*	*ütelda*
what:NOM	there	else:PRT	say:INF

'What else [was] there to say.'

d. CED (Coastal)

ei	*minul*	*ole*	*mieles*	***mitta***
not	me:ADE	be:CNG	mind:INE	what:PRT

eit	*rääkkis*
old_woman	say:PST:3SG

'I don't remember what the old woman said.'

e. PCESS

ja	*siis*	*nemad*	*et*	*siis*	*kiiresti*	*kodumaale*
and	then	they	so_that	then	quickly	homeland:ALL

saada	***mida***	*nad*	*tegid*	*siis*	*Taišeti*
get:INF	what:PRT	they	do:PST:3PL	then	Tayshet:GEN

jaamas	*muidu*	*me*	*käisime*	*Rutiga*
station:INE	by_the_way	we	go:PST:1PL	Rutt:COM

turismirongiga	*nendes*	*paikades*	*ka*
tourist_train:COM	this:PL:INE	place:PL:INE	too

'And then they, to get quickly [back] to the homeland, what they did then in Tayshet station, by the way Rutt and I also visited those places on a tourist train.'

Polarity of the main verb. This marks whether the main verb of the clause that includes the pronoun *mis* is affirmative or negative (5). As mentioned previously, if a sentence's polarity is negative, the object in that sentence is typically always partial, regardless of other syntactic factors (Metslang 2017: 267). That means it may be more likely that the pronoun *mis* is in partitive when the verb in the clause is negated.

(5)   CED (Võru)

*tuud   ei   võiq   kõnõlda   **midä**   ei   tiijä*
that:PRT   not   can:CNG   speak:INF   what:PRT   not   know:CNG
'[One] cannot talk about what [one] does not know of.'

Tense of the main verb. This marks whether the main verb of the clause that includes the pronoun *mis* is in the present tense (6a), the past tense (this also includes the perfect and pluperfect tenses; 6b) or occurs as an independent infinitive (6c).

(6)   a.   PCESS

    *see   on   üks   küsimus   **mida**   tahaks*
    this   be:3SG   one   question   what:PRT   want:COND
    alguses   enda   käest   küsida
    beginning:INE   myself:GEN   from   ask:INF
    'That is one question that [I] would first like to ask myself.'

  b.   CED (Insular)

    *miä   mälettä   ühti   **miss**   siält*
    I   remember:CNG   at_all   what:NOM   from_there
    maksõtti
    pay:IPS:PST
    'I do not remember at all what they paid there.'

  c.   PCESS

    *see   on   viimane   asi   **mida**   soovida*
    this   be:3SG   last   thing   what:PRT   wish:INF
    'That is the last thing to wish for.'

Object type of the main verb. This marks whether the main verb of the clause that includes the pronoun *mis* is a partitive verb (7) or a transitive verb that allows systematic object case alternation. Verb classification is one of the main things that object case alternation in Estonian is dependent on: certain verbs, called partitive verbs, only allow partitive partial objects, as opposed to other verbs, called aspect verbs, which allow the alternation of object case (Erelt et al. 1993: 49, Erelt 2017: 113–114). Aspect verbs (e.g. *koostama* 'compile', *äratama* 'wake', *võtma* 'take', *kutsuma* 'call, invite', *tooma* 'bring') have a boundary or an endpoint of a situation in their lexical meaning, which

allows the opposition of total and partial objects, while partitive verbs (e.g. *armastama* 'love', *huvitama* 'interest', *üritama* 'try', *tähendama* 'mean') cannot express a temporal boundary and are consequently confined to partitive objects (Tamm & Vaiss 2019: 160). Therefore, if a verb only allows partial objects, then it is likely that with that verb the partitive *mida* is used more frequently than the nominative *mis*.

The classification of partitive verbs has been somewhat debatable in Estonian linguistics, as some verbs that are considered to be partitive verbs by Erelt et al. (1993) can still occur with total objects if there is an element in the sentence that adds a result or boundary to the situation. In opposition, some linguists have classified partitive verbs into so-called "hard" and "soft" partitive verbs; the former truly only allow partitive partial objects, while the latter can also occur with total objects in nominative or genitive if a perfective element has been added to the sentence (Rätsep 1978: 221–223, Klaas 1999: 55, see also Tamm 2003, Tamm & Vaiss 2019). In this study, the set of partitive verbs was chosen to include only the "hard" partitive verbs, while the "soft" partitive verbs have been categorised as other verbs. This sorting is based mostly on the classification in the object dictionary[7], which distinguishes the two partitive verb types by presenting alternative object case options for those verbs that allow total objects, if certain elements are added to the sentence. The sorting has been done on the basis of Standard Estonian and it is possible that there are slight differences in the use of the verbs regarding object cases in the dialects.

(7)   PCESS
     projekt   ***mida***   *üritame*   *kirjutada*   *praegu*
     project   what:PRT   try:1PL   write:INF   currently
     'A project that we are currently trying to write.'

---

7   http://www.eki.ee/dict/asp/ (Accessed December 17, 2020.) The object dictionary has been compiled primarily for L2-learners of Estonian and it consists of the 400 most frequent transitive verbs in Estonian. The dictionary provides information about which verbs require an object in the partitive case and which verbs also allow objects to be in nominative and/or genitive. However, due to the fact that the main aim of this dictionary is to assist L2-learners, it might not include all marginal uses of object cases for some verbs.

Use of impersonal voice in the clause. This marks whether the main verb of the clause that includes the pronoun *mis* is personal (active) or impersonal (passive, 8). Impersonal voice retains the opposition of total and partial objects, but total objects occur in nominative instead of genitive, which means that total objects in impersonal sentences have some characteristics of subjects (Erelt 2017: 210, 217, see also Nurme 2012). Impersonal voice is also one of the contexts where it is obligatory to use partitive for personal and reflexive pronouns (see examples 2a–b), even when according to the general DOM rules they should be total objects (Metslang 2017: 272–273).

(8) CED (Insular)

nee	*peenemad*	*toidud*	***miss***	*nüid*	*tehtakse*
this:PL	fancy:CMP:PL	food:PL	what:NOM	now	make:IPS

'These fancier foods that are made nowadays.'

Length of the following word. This marks whether the word immediately following the pronoun *mis* has one, two or more than two syllables (9a–c). Monosyllabic words are, furthermore, categorised into pronouns and other parts of speech. The category of pronouns includes short personal pronouns (*ma* 'I', *sa* 'you', *ta* 's/he', *me* 'we', *te* 'you (pl)', *nad* 'they'), demonstrative pronouns *see* 'it' and *need* 'these', and their monosyllabic case forms (in the CED dataset, the dialectal equivalents *too*, *taa* 'it' and *nood*, *naad* 'these' were also marked). These monosyllabic words, together with the pronoun *mis*, tend to form one foot (i.e., a speech unit from one accent to the next, which includes one stressed syllable, followed by one or two unstressed syllables (Asu et al. 2016: 126–127)). This might increase the likelihood of the partitive *mida* to be shortened into the nominative *mis*. Therefore, it is possible that the shorter the word following the pronoun *mis*, the higher the chance that the nominative case is used. The reason behind separating the personal and demonstrative pronouns from other monosyllabic words is two-fold. Firstly, these pronouns are often in the position of the subject and, therefore, could influence the choice between *mis* and *mida*. Secondly, the constructions of *mis* + PRONOUN are very frequent in Estonian (see Raudvere & Uiboaed 2018), and since high-frequency constructions are much more likely to undergo change (see e.g. Bybee 2010, Diessel 2017), it is possible that this particular linguistic context encourages the use of the nominative *mis*.

(9)   a.   PCESS

> *ma*   *ei*   *taha*   *mõeldagi*   **mis**   *see*
> I   not   want:CNG   think:INF:CLI   what:NOM   it
> *rehvidega*   *teeb*
> tire:PL:COM   do:3SG
> 'I don't even want to think what that does to the tires.'

b.   CED (Seto)

> **midä**   *lövvä*   *tuud*   *võtta*
> what:PRT   find:1SG   this:PRT   take:1SG
> 'I take what I can find.'

c.   PCESS

> *see*   **mida**   *mõistetakse*   *muusika*   *all*   *võib*
> this   what:PRT   consider:IPS   music   under   can
> *olla*   *kultuuriti*   *sootuks*   *erinev*
> be:INF   culture:DISTR   utterly   different
> 'What is considered to be music can be utterly different across cultures.'

Dialect. This marks which dialect area the speaker is from: the Eastern, Mid, Western, Insular, Coastal, Northeastern, Mulgi, Tartu, Võru or Seto dialect. This variable is most useful to discern any possible impact of language contact, since different dialectal areas have been affected by different contact languages over time. The variable is only applicable in the CED dataset.

A final note on the variables: although there are a few variables included in this study that, according to Standard Estonian rules, should have values that only allow partial objects in partitive (i.e., verb type: partitive verb, polarity: negative), this is not the case for the datasets used in this study. This means that for every value of each variable, there are instances of both the nominative *mis* and the partitive *mida* in both datasets.

Table 3. The coded variables and their possible values. If applicable, the abbreviations used in subsequent graphs are given in parentheses.

Variable	Values
case marking of *mis* (CASE)	nominative (nom) partitive (prt)
clause type (CLAUSE_TYPE)	relative clause (rel), relative clause without a main clause (rel_main) question (q), indirect question (indir_q) rhetorical question (rhet)
polarity of the verb (POLARITY)	affirmative (aff) negative (neg)
tense of the verb (TENSE)	present (pr) past (pst) infinitive (inf)
verb type (VERB_TYPE)	partitive verb (prt) other verb (other)
impersonal voice (IMPS)	yes no
following word length (SYLLABLE)	one syllable, pronoun (1_pron) one syllable, other part of speech (1_other) two syllables (2) more than two syllables (3+)
dialect	Eastern (EST) Western (WST) Mid (MID) Insular (INS) Coastal (CST) Northeastern (NE) Mulgi (MUL) Tartu (TRT) Võru (VRU) Seto (STO)

2.3. Methods

To determine which variables most influence the variation between *mis* and *mida* in partial object position, I have analysed both datasets using three different methods: two non-parametric tree-based models – conditional inference trees and random forests (Breiman 2001, Strobl, Malley & Tutz 2009) – and a classification method of binomial mixed-effects logistic regression (Baayen, Davidson & Bates 2008).

Logistic regression has been used in quantitative variation studies for a while, but it is a more recent approach to apply mixed-effects models, random forests and conditional inference trees to the variation data. Furthermore, studies like Tagliamonte and Baayen (2012) and Baayen et al. (2013) have shown that it is beneficial to use all three of these complementary techniques together. Mixed-effects models provide a way to account for random-effect factors (e.g. characteristics of the speakers in the data). Random forests, in turn, provide information about the importance of predictors. Finally, conditional inference trees visualise how different predictors operate in association with each other.

The goal of all of these methods is the same: to predict the chances of the dependent variable (in this study, to predict the chances of a partitive *mida* in partial object position) occurring in a given context specified by the other variables. These methods have also been successfully used in other studies of Estonian and dialect syntax (see e.g. Klavan, Pilvik & Uiboaed 2015, Reile 2015, Ruutma et al. 2016, Lindström & Uiboaed 2017, Taremaa 2017, Lindström, Pilvik & Plado 2018, Pook 2019, see also Basile & Ivaska 2021, Klavan 2021, and Hint et al. 2021, this volume).

All the calculations were performed using the statistical software R (R Core Team 2018). The conditional inference trees and random forests were computed using the functions *ctree()* and *cforest()* from the package *party* (Hothorn, Hornik & Zeileis 2006). The binomial mixed-effects logistic regression was computed using the function *glmer()* from the package *lme4* (Bates et al. 2015).

3. Results

In this section I present and analyse the results from the data: Section 3.1 contains the analysis for the spontaneous speech dataset and Section 3.2 contains the analysis for the dialect dataset. Both datasets have been analysed with two of the aforementioned methods – the conditional inference tree model and the random forest model. In Section 3.3 I combine the two datasets into one, use binomial logistic regression, compare the results obtained from the different analyses and datasets, and explain the possible reasons behind the significance of the variables and their values.

3.1. The analysis and results of the spontaneous speech data

To analyse the way the variables interact with each other in the choice between *mis* and *mida* in partial object position, I first ran a conditional inference tree model. The conditional inference tree is a method based on binary recursive partitioning, wherein at each stage the algorithm tests whether any of the independent variables are significantly associated with the given response variable. The variable that is most strongly associated with the response is selected for the next split, which divides the dataset into two subsets. This continues until there are no variables that are associated with the response at the level of statistical significance. The result of this process is depicted as a tree structure with binary splits. Therefore, conditional inference trees can show us which of the variables influence the choice between *mis* and *mida* most significantly and how the variables are associated in relation to each other.

Figure 2 shows the conditional inference tree graph for the case of the pronoun *mis* in partial object position in spontaneous speech. The variables in the model are clause type, polarity, the use of impersonal voice, tense, verb type and length of the following word. The splits are made in each node by the variable that is most strongly associated with the response: the case of the pronoun *mis*. The graph presents all the possible splits significant at the level of 0.05 or less. The bar plots at the bottom show the proportions of nominative (light grey) and partitive (dark grey) pronouns in each end node; together, these contain all the observations with the given combination of features.
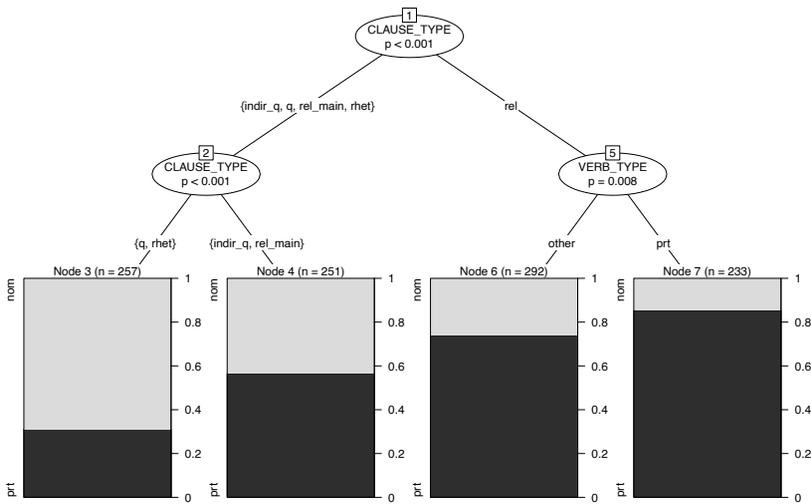
Figure 2. Conditional inference tree for the case of the pronoun *mis* in partial object position in spontaneous speech.

As can be seen from Figure 2, the case of the pronoun in spontaneous speech is mainly dependent on clause type: in relative clauses the use of *mida* is higher than in other clauses. Relative clauses are next split by verb type: if the verb of the clause is a partitive verb, then the chances of *mida* occurring in partial object position are very high (Node 7; example 10); for any other verb the probability is lower, but still relatively high (Node 6).

(10) PCESS

sa	*teed*	*täpselt*	*seda*	***mida***	*su*
you	do:2SG	exactly	this:PRT	what:PRT	you:GEN

vanemad	*sult*	*ootavad*
parent:PL	you:ABL	expect:3PL

'You are doing exactly what your parents expect of you.'

The rest of the clause types are split again by clause type, separating questions and rhetorical questions from the other types; for these two types, the probability of using partitive in partial object position is significantly lower (Node 3) than in the other clause types (Node 4).

In order to assess the goodness-of-fit of this model (and all the following models), I have calculated the C-index of concordance (also known as the area under the ROC-curve), which is the proportion of

concordant pairs divided by the total number of possible evaluation pairs and which evaluates the predictions made by the algorithm. A value of 0.5 means that the model is not able to discriminate between the variants, while a value over 0.9 means that the model can discriminate between the dependent variable values outstandingly well. The C-index of this tree is 0.73, which indicates a moderately good fit.

The conditional inference tree shows the associations between significant variables, but does not indicate which of those significant variables have the highest impact. Therefore, I ran a random forest model, which outputs the measure of importance for every variable included in the model, averaged over many conditional inference trees (in this study, 1000 trees). These (permutation) variable importance measures reflect the impact value that each predictor has on the response, i.e., the dependent variable, which in this study is the case of the pronoun *mis*. The conditional importance scheme has been used for the computation of these measures in order to take into account the possible correlations between the variables and help identify the truly relevant predictor variables (see Strobl et al. 2008).

This analysis includes all the same variables that were in the conditional inference tree model. The impact of the variables is presented in Figure 3. The names on the y-axis indicate the variables included in the analysis. The numbers on the x-axis represent the relative difference between the importance of the variables.
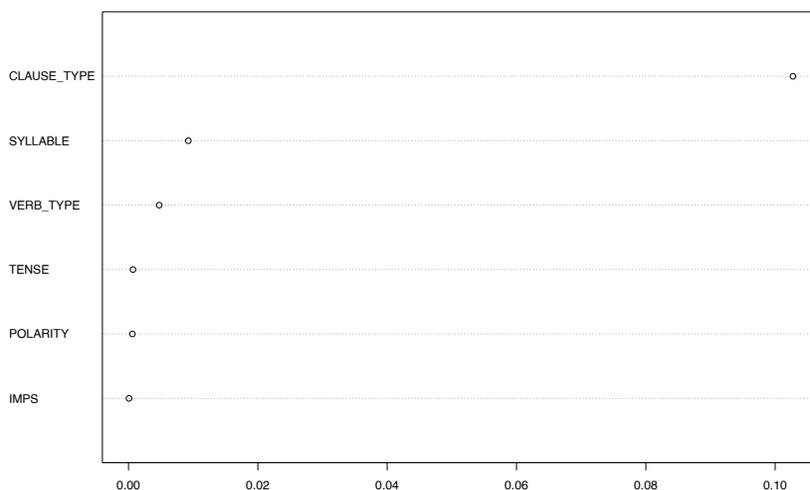

Figure 3. Random forest for the case of the pronoun *mis* in partial object position in spontaneous speech.

We can see from Figure 3 that the most important predictor for the case of the pronoun *mis* in partial object position is clause type (0.105), followed by the length of the following word (0.009) and verb type (0.005), while the rest of the variables (tense, polarity and impersonal voice) do not seem to have much or any discriminatory power. The C-index of concordance for the model is 0.79, which indicates a moderately good fit.

3.2. The analysis and results of the dialect data

In order to analyse the variation of *mis* and *mida* in dialects, I again started with the conditional inference tree model. Figure 4 displays the conditional inference tree graph for the case of the pronoun *mis* in partial object position in the dialect data. The variables in the model are dialect, clause type, polarity, the use of impersonal voice, verb type, tense and the length of the following word, the response is the case of the pronoun *mis*. The graph can be interpreted in the same way as the graph in the previous section.
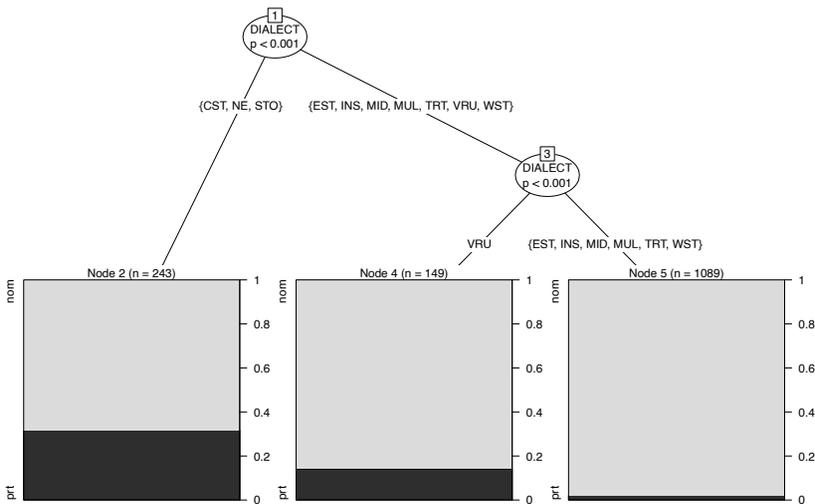
Figure 4. Conditional inference tree for the case of the pronoun *mis* in partial object position in dialects.

Figure 4 shows that in the dialects the most important and indeed the only predictor (in this model) affecting the choice between *mis* and *mida* in partial object position is the dialect itself. This first divides the dataset into two: in the Northeastern, Coastal and Seto dialect, the possibility of using a partitive pronoun is higher (Node 2) than in the rest of the dialects (Node 3), which are split again by dialect, separating the Võru dialect from the Eastern, Mid, Western, Mulgi, Insular and Tartu dialects; the former also has a higher percentage of partitive pronouns in partial object position than the other dialects. The C-index of this tree is 0.82, which indicates a good fit.

However, in order to assess whether there are any other variables that might significantly affect this variation in the dialects, I removed the variable of dialect and ran the model again. Figure 5 displays this modified conditional inference tree graph.
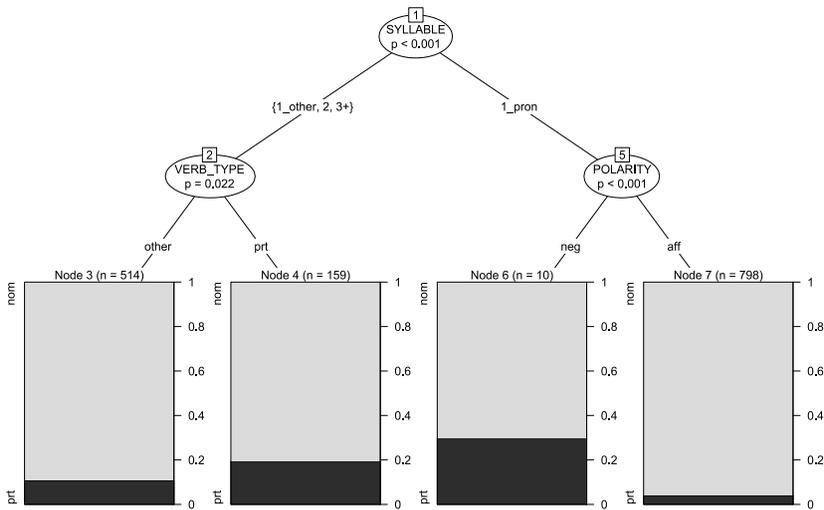
Figure 5. Conditional inference tree for the case of the pronoun *mis* in partial object position in dialects, the variable of dialect excluded.

As can be seen from Figure 5, after removing the variable of dialect, three other variables appear significant in determining the choice between *mis* and *mida*. The most important predictor is the length of the following word, which separates the monosyllabic pronouns from other following word length values. For the monosyllabic pronouns, the next split is made by polarity, as it is more likely to use a partitive *mida* in negative clauses (Node 6) than in affirmative clauses (Node 7). The

other word length values are next divided by verb type: the likelihood of a partitive *mida* is higher with partitive verbs (Node 4) than with other verbs (Node 3). The C-index of this tree is only 0.67 (compared to the previous model's 0.82), which indicates that the previous model is able to discriminate between the dependent variable values much better than this one. Therefore, the dialects do play an important role in the choice between the case forms of the pronoun *mis*.

Next, the random forest model was applied in order to assess the importance of the variables included in the analysis, which are the same as they were for the conditional inference tree model (now again including the variable of dialect). The impact of the variables in the analysis is presented in Figure 6. The graph can be interpreted in the same way as the graph in the previous section.
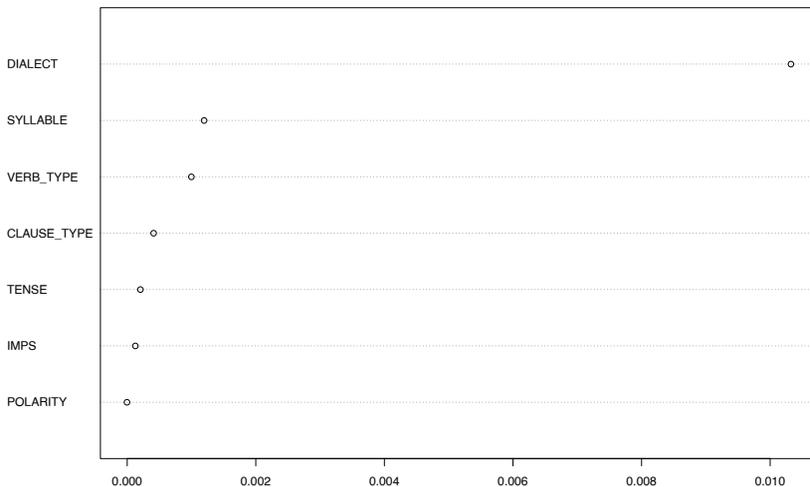
Figure 6. Random forest for the case of the pronoun *mis* in partial object position in dialects.

From Figure 6 we can see that the most important determiner of the case of the pronoun *mis* in the dialects is the dialect itself (0.010), followed by verb type (0.001), length of the following word (0.001) and clause type (0.001). Tense, impersonal voice and polarity do not seem to have much or any discriminatory power. The C-index of concordance for the model is 0.92, indicating a great fit.

3.3. Combined analysis and comparison of the datasets

As a final analysis method, I used binomial mixed-effects logistic regression. This method is suitable for explaining and predicting speakers' choices between two variants on the basis of geographic, social, syntactic, pragmatic and other factors. A mixed-effects model includes both fixed effects, whose effect is measured, and random effects, which are sampled randomly from the population (e.g. individual speakers, lexical stimuli, individual lemmas, etc.) and that are useful in filtering out "noise" from the model. Logistic regression takes a predetermined configuration of predictors as a reference and compares the change in effect on the response (which is the case of the pronoun *mis*) for each predictor value, while keeping other predictors constant.

While the methods used above, namely conditional inference trees and random forests, do not set any distributional requirements for the data and are, therefore, suitable to use in the case of unbalanced data, logistic regression models are more particular about the distribution of the data points. As is often the case with dialectal data, my dialect dataset does not meet all the requirements for using logistic regression, because half of the dialects have less than five instances of *mida*. Consequently, in order to still apply this method to my data, I have combined the two datasets into one. This new dataset has an additional binary variable called *data*, which shows whether an observation comes from the spontaneous speech dataset (PCESS) or the dialect dataset (CED).

The new dataset also omits the variable of dialect. Oftentimes it is not acceptable to combine all the dialects into one big "dialect" of the Estonian language, since the dialects are typically not homogenous. However, in this case, we have already seen that there is very little variation between the dialects in terms of the use of the pronoun *mis* in partial object position: most dialects use the nominative *mis* in almost all instances. Therefore, we can continue with this combined dataset, but just have to keep in mind while interpreting the results that there are small dialectal differences not represented in the model output.

For this model I again included the same variables as for the conditional inference tree and the random forest models. In order to potentially confirm any significant influence of the speakers (see e.g. Van de Velde & van Hout 1998, Tagliamonte & Baayen 2012), I have added them into the regression model as a random effect.

Comparing the Akaike information criterion (AIC) values of models containing different variables and variable interactions, I found that the most adequate model for predicting the case of the pronoun *mis* in this combined dataset is CASE ~ CLAUSE_TYPE + POLARITY + VERB_TYPE + IMPS + SYLLABLE + DATA + CLAUSE_TYPE *DATA + VERB_TYPE*IMPS + 1|SPEAKER.

The reference configuration is a context where the clause type is a relative clause, the clause is negative and the verb is used in the personal/active voice, the verb type is a partitive verb, the word following the pronoun *mis* has more than two syllables and the data comes from the spontaneous speech dataset (i.e., a context in which the use of *mida* is most likely based on the previous two models and contingency tables of the data; 11).

(11) PCESS

selline	*saast*	***mida***	*omanik*	*ei*	*taha*
this_kind_of	crap	what:PRT	owner	not	want:CNG

'The kind of crap that the owner doesn't want.'

The intercept value is 3.64, which means that the chances of the pronoun *mis* being in partitive in this configuration are 38.04 times higher than those of it being in nominative, and the probability of a partitive pronoun in this context is 0.97. Therefore, in this chosen context it is very likely that the speaker chooses a partitive pronoun over a nominative pronoun in partial object position.

Figure 7 illustrates the trends in the logistic regression model for each variable. In contrast to the aforementioned variable values that increase the likelihood of the partitive *mida*, we can see from the figure that the probability of using *mida* decreases when the clause is affirmative and the pronoun is followed by a monosyllabic pronoun. The interaction between clause type and data shows that for the dialects, using *mida* is relatively unlikely in all clause types, but in the spontaneous speech dataset, the probability of using partitive increases in relative clauses and decreases in (rhetorical) questions. The interaction between verb type and the use of impersonal voice shows that, for partitive verbs, it is more likely to use *mida* in personal voice and *mis* in impersonal voice, while for other verbs, the relation is the opposite. However, the 95% confidence interval for impersonal partitive verbs is far too large

to make any clear-cut conclusions about the significance of this interaction. The C-index of this model is 0.88, which indicates a great fit.
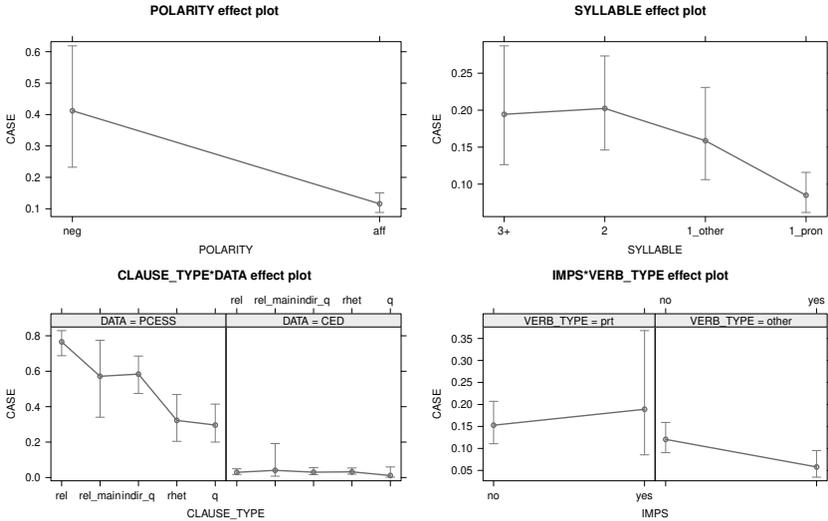

Figure 7. The logistic regression predictor effects for predicting the case of the pronoun *mis* in partial object position in the combined dataset. The bars for each predictor value depict the 95% confidence intervals.

The range of variation in the use of *mis* and *mida* in spontaneous speech and in the dialects differs significantly, so it is clear that these two datasets and their models are not entirely comparable. However, taking into account the variables which appeared significant with (almost) all of the three used methods, it can be generalised that the observable variation is mostly affected by **verb type**, **clause type** and **the length of the following word** (and of course **dialect** for the dialect data). Although these variables influence the choice between *mis* and *mida* to varying degrees in both datasets and even in the different models of the same dataset, they can all still be considered important in describing the variation.

The reasons behind the significance of some of the variables are quite clear. For example, a partitive verb typically requires a partitive partial object in simple sentences, so it can only be expected that this variable influences the choice between *mis* and *mida* in partial object position in a similar way. It is interesting to note here that, although the verbs were classified into partitive and other verbs according to the

strict rules (i.e., the partitive verbs in the dataset are the "hard" partitive verbs, which should only allow a partitive partial object), there are still plenty of clauses in the data where those hard partitive verbs allow a nominative *mis* in partial object position. At first one might think that this is due to the fact that the verb sorting was based on a learners' dictionary, which may have omitted marginal uses of object cases for some verbs. While this is certainly true for several verbs, there are also examples that include verbs that have been previously distinguished as genuinely "hard" partitives (see e.g. Tamm & Vaiss 2019: 173), like *kartma* 'fear', *mõjutama* 'influence', *mäletama* 'remember', etc. (12a–b). That means that the already quite debatable classification of partitive verbs has been made even more problematic by this analysis; even when applying a strict approach, the "hard" partitive verbs allow a nominative object, although possibly only in the case of the pronoun *mis*.

(12) a. PCESS

mis	*siin*	*ikka*	*karta*
what:NOM	here	ever	fear:INF

'What is there to fear?'

   b. CED (Eastern)

ma	*tahan*	*riakki*	***mis***	*ma*	*viel*
I	want:1SG	talk:INF	what:NOM	I	still

melettan
remember:1SG

'I want to talk about what I still remember.'

The significance of the length of the following word may be related to prosodic structure of a sentence – since the pronoun *mis* followed by a monosyllabic word forms one foot, this might encourage the partitive *mida* to be shortened into the nominative *mis*. In addition, the analyses showed that there is an even higher likelihood of *mis* when the word following is a monosyllabic (personal) pronoun, which can most likely be attributed to the high usage frequency of such pronouns. Word frequency and its influence on this variation is discussed in more detail in Section 4.2. Furthermore, in the constructions MIS + SA and MIS + SEE, the assimilation of *s* could be another factor in the speakers' choice of preferring the nominative *mis*.

For both datasets, questions and rhetorical questions are distinct in their infrequent use of the partitive *mida* in partial object position: in the spontaneous speech data, both question types have an average of 30.3% of *mida*, while all the other clause types have an average of 65.4% of *mida*; the same figures for the dialect data are 4.4% and 10.2%, respectively. It is not quite clear why these two clause types differ from others (see examples for the different clause types in Section 2.2).

For *wh*-questions and indirect questions, it has been said that the focus (and possibly the stress) is on the question word, but in the case of relative clauses, the relative pronoun is the topic instead, and the focus is somewhere else in the sentence (Pajupuu 1990: 147, Lindström 2006: 881–883). Additionally, Laanesoo (2014: 121–122) has found that in *mis*- or *mida*-questions, sentence stress is on the verb; however, in the case of directives (i.e., questions that were meant as directives and not as requests for information) starting with *mida*, the stress is both on the question word and the verb.

Assuming that the correlation between sentence stress and word length that has been proven to exist for personal pronouns (i.e., that long personal pronouns are in a more stressed position than short personal pronouns (see Pool 1999)) also applies to interrogative-relative pronouns, it would mean that in questions the pronoun *mis* should be stressed and is therefore longer, i.e., the partitive *mida*. That, however, is not the case.

One explanation could stem from the words following the pronoun *mis*: more often than not the interrogative pronoun *mis* in questions is followed by a monosyllabic personal or demonstrative pronoun. On average, 75.1% of the interrogative pronouns in the spontaneous speech data and 72.4% of the interrogative pronouns in the dialect data are followed by a short pronoun, while for other clause types the same figures for (relative) pronouns are 54.2% and 50.5%, respectively. Since the analysis showed that a following short pronoun increases the likelihood of nominative *mis*, it could be possible that the way (rhetorical) questions are typically formed, including a short pronoun, is the reason why they differ that much from other clause types. However, since the interaction of clause type and the length of the following word was not actually statistically significant in the logistic regression model, the reason behind the contrast between questions and other clause types might lie elsewhere. For example, it could also be related to the interrogative

286 Hanna Pook

pronoun *mis* being sentence-initial, since in all of the other clause types the pronoun is sentence-internal.

The variable of dialect is especially interesting in regard to how it divides the dialects. Typically Estonian dialects are divided into southern and northern dialects, based on the differences found in their phonology, morphology and lexis. In this case, however, a distinction can be drawn between the eastern (i.e., the Northeastern, Coastal, Seto and Võru) and western dialects instead. This division of the dialects into eastern and western instead of northern and southern is, nonetheless, not at all unusual, as recent studies in dialect syntax have attained similar results (see e.g. Lindström et al. 2009, Uiboaed 2013, Uiboaed et al. 2013, Lindström, Uiboaed & Vihman 2014, Lindström et al. 2015, Lindström, Pilvik & Plado 2018).

Nevertheless, individual differences between speakers cannot be disregarded. The choice between *mis* and *mida* is definitely related to and affected by both syntactic and geographical factors, but also varies between the speakers, and at times, it varies quite a lot. For example, there are speakers even from the Coastal, Northeastern and Seto dialects who exclusively use *mis* in partial object position, despite the overall percentage of *mida* being higher in those dialects than in other dialects. Therefore, it would be more accurate to make conclusions about the language of the subdialects rather than about entire dialects, since the use of the pronoun *mis* in the language of one subdialect's speakers is often (but not always) more uniform. The percentage of the partitive pronoun *mida* in partial object position by subdialects is presented in Figure 8.

According to Figure 8, it seems that the deductions about the Seto and Northeastern dialects are mostly accurate, since almost all of their subdialects (except Iisaku in the Northeastern dialect) use a fair amount of *mida* in partial object position, but the same cannot be said for the Coastal dialect, where virtually only one subdialect (Viru-Nigula) raises the percentage of the whole dialect. It could be possible that the use of *mida* in that one subdialect is connected to bordering and therefore having close contacts with the Northeastern dialect, but in general this is not a characteristic used to describe the Coastal dialect itself.
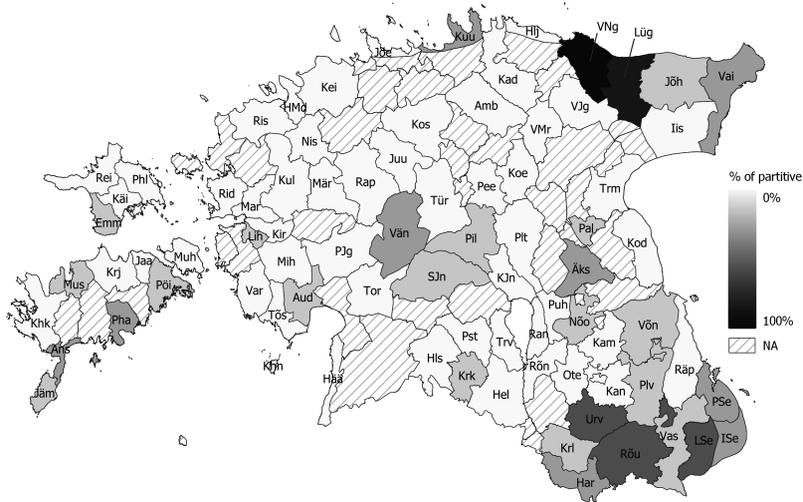
Figure 8. The percentage of the partitive pronoun *mida* in partial object position in the subdialects represented in the data. It should be noted that if data from a subdialect is not available (NA), it does not necessarily mean that the subdialect is not included in the corpus, but rather that the conversations from these subdialects simply did not have any pronouns *mis* in partial object position.

Oftentimes, however, a subdialect is only represented by a few speakers who, in turn, can contribute very few observations to the data. Taking that into account, one should not view the map in Figure 8 as the complete truth, but more as an illustration of the tendencies concerning this particular variation.

4. Motivations behind the variation

In this section, I address three possible explanations for the emergence and development of the *mis ~ mida* alternation in partial object position: language contact, high usage frequency and language standardisation; these are discussed in Sections 4.1, 4.2 and 4.3, respectively.

4.1. Language contact

The first possible explanation for the case variation of *mis* is language contact, which has also been speculated to be behind this variation by other linguists. Lindström, Uiboaed and Vihman (2014: 617–618) have suggested that Estonian has copied this variation from Russian case syncretism, since Russian does not differentiate between nominative and accusative *что* 'what' when referring to inanimate masculine and neuter words. Similarly, Kont (1963: 109) believes that it is possible that the general structure of Russian objects has influenced the similar variation in Veps. In the Risti subdialect, Juhkam (1983: 122) considers it to be the result of Swedish influence, since Swedish has no opposition of object cases.

However, there have also been claims against the influence of contact languages: for Veps, Kettunen (1943: 107) argues that contact with Russian alone could not have caused this variation in partial object position. Instead, he explains that due to apocope (the loss of a word-final vowel), the form for the partitive singular became identical to the form for the nominative plural in all vocalic-stem nouns (e.g. *poigad* 'son:PRT, son:PL', *mad* 'land:PRT, land:PL'), and in order to avoid the confusion in meaning, an object case other than partitive came into use.

Oftentimes it is possible to examine language contact in the language of the dialects. While it can be difficult to make any categorical conclusions in the case of morphosyntactic variation, the statistical frequency of one variant in certain geographical areas (i.e., in certain dialects) could indicate the influence of certain language contacts (see Koptjevskaja-Tamm & Wälchli 2001: 627, Kortmann 2010). Meanwhile, the same information would be unattainable from the standard spoken language, which is relatively uniform across all speakers, irrespective of where they live in Estonia.

The results of the dialect dataset showed that using the nominative *mis* in partial object position is prevalent in all the dialects, but the eastern dialects, or more specifically certain subdialects in the eastern dialects, still have retained some possibility of using the partitive *mida* in the same position.

This result seems to indicate that Russian influence might not be the reason for the case variation of *mis* in partial object position as Lindström, Uiboaed and Vihman (2014) have speculated. In that case,

the eastern dialects, which have had closer contact with Russian than other dialects, should exhibit a clearer preference for the nominative *mis*. Instead, the eastern dialects are the only ones which have evident case variation in this position.

The influence of Swedish, however, which Juhkam (1983: 122) has speculated to be the reason behind using total object cases for partial objects in the Risti subdialect, could apply in the case of the Insular and Western dialects. The speakers of these dialect areas have had the most contacts with the Swedish (Ariste 1939: 5, Juhkam & Sepp 2000: 58, Pajusalu et al. 2009: 123) and in the data of this study, they do exhibit few or no partitive uses in partial object position. Nevertheless, Swedish influence would not explain the prevalence of the nominative *mis* in the Mid, Eastern and Mulgi dialects. Therefore, it is possible that the absence of the partitive *mida* in most dialects is the result of something other than late local contacts.

Still, language contact in general, or maybe rather the lack of it in certain areas, cannot be disregarded as a factor in this variation. As mentioned before, the pronoun *kes* 'who' does not exhibit a similar variation in partial object position, but is used according to the general DOM rules. This kind of contrast between the animate and inanimate interrogative and/or relative pronoun is, however, not uncommon in Indo-European languages. If, in addition, we take into account that the north- and southeastern dialects have, over time, been much more conservative in terms of change and have not been affected as much by Indo-European influences as the rest of the dialects (see e.g. Keem & Käsi 2002: 22, Pajusalu et al. 2009: 64, 82), it is possible that the lack of a general Indo-European or specifically Germanic impact in certain areas (which is, according to previous studies, often difficult to determine (see Lindström et al. 2019: 159–160)) could explain the differences between the dialects.

4.2. High usage frequency

Another explanation for the development of this variation could be high usage frequency. Frequency plays an important role in the usage-based analysis of language since high-frequency words or constructions are much more likely to undergo change or to change at a faster rate than low-frequency words or constructions (see e.g. Bybee 2010,

Diessel 2017). The pronoun *mis* is not only the most frequent of the interrogative-relative pronouns, but is also among the most frequent words in the Estonian language in general (see Kaalep & Muischnek 2002, Raudvere & Uiboaed 2018). Therefore, it is in a favourable position to undergo change.

This is typically a change towards brevity, as previous studies have found that frequently used words or constructions also tend to be shorter and are oftentimes phonetically reduced (Tauli 1968: 30, see also Hooper 1976, Bybee 2001, Diessel 2007). A psycholinguistic explanation for this correlation has been given by Jurafsky et al. (2001): speakers want to produce speech with minimal effort, but at the same time they must ensure that the hearer is able to understand the intended meaning. However, since frequent words are more expected or predictable in a conversation, they are most likely to be understood even in their reduced form, and therefore, speakers can expend only a minimal effort in producing such words (see also Gregory et al. 1999, Bybee 2010).

If, in addition, individuals have stopped perceiving the difference in meaning of certain case forms, i.e., between the nominative *mis* and the partitive *mida* in partial object position (especially since the grammatical role of *mis*/*mida* as a partial object can often be interpreted from the transitive verb and/or the main clause), it is very likely for them to prefer the shorter form over the longer form.

It must be acknowledged that frequency can also have the opposite effect of making words and word forms more resistant to change (see e.g. Hooper 1976, Bybee & Slobin 1982, Bybee 1985, Ogura 1993, Bybee 1995). A good example of this are the irregular past tense forms in English. However, in the case of the variation between *mis* and *mida*, frequency has likely rather encouraged change than hindered it.

The correlation between frequency and (phonetic) reduction has been shown to be especially strong in linguistic sequences: oftentimes a frequent collocation conditions a reduction which would otherwise not occur in individual words or in a less frequent combination (see Krug 1998, Bybee & Scheibman 1999, Jurafsky et al. 2001, Krug 2003).

In order to examine whether the frequency of the pronoun *mis* or that of constructions involving the pronoun has played a role in the development of this variation, I included the distinction of monosyllabic pronouns in the following word length variable. The analysis showed that if the pronoun *mis* was followed by a monosyllabic personal or

demonstrative pronoun, the likelihood of using the nominative *mis* was higher than with other monosyllabic or longer words.

The significance of this can be related specifically to frequency: the construction of MIS + MONOSYLLABIC PRONOUN is one of the most frequent amidst the MIS + X constructions in Estonian: based on newer bigram frequency lists compiled by Raudvere and Uiboaed (2018) on newer fiction texts, out of the five most frequent MIS + X (lemma) constructions, four are MIS + (SHORT) PRONOUN. Therefore, the frequency of this construction can be the reason why it has a higher percentage of nominative *mis* than other constructions.

4.3. Language standardisation

Lastly, this variation could also be influenced by language standardisation. Language planning in Estonia is considered to have first begun in the 1870s, but started to really thrive in the beginning of the 20th century (Erelt, Erelt & Ross 2007: 12–14). Although the variation of *mis* and *mida* in partial object position has never been, to my knowledge, standardised in any way, it is clear that the partitive *mida* in this position follows the general rules of DOM, while using the nominative *mis* is an exception to the rule.

Therefore, the distinction between total and partial objects should be more evident in the spoken variant of the standard language than in the dialects, which are based solely on oral tradition and have had little to no influence from standardisation. Moreover, the speakers in the dialect corpus have received their (often only a few years' worth of) education at a time when the Estonian language was mostly still unstandardised. In addition, German and even some Estonian authors of that time did not always use object cases in a way similar to the current rules (Metslang & Habicht 2019), which might also have influenced the speakers' use of the pronoun *mis*.

When comparing the two datasets included in this study, we can see that the range of variation concerning *mis* and *mida* in spontaneous speech and dialects is distinctly different: in the spontaneous speech data, the use of nominative and partitive in partial object position is more or less equal, while in the dialect data, the partitive *mida* occurs in less than 10% of the sentences (see Table 2).

Let us now add another language register to this comparison – fiction texts – and compare the use of the previously analysed construction of *MIS* + MONOSYLLABIC PRONOUN in them. In the previous section I looked at the lemma bigram constructions of *MIS* + X, in which the constructions of *MIS* + MONOSYLLABIC PRONOUN stood out as being at the top of the frequency list. Here, however, I examined the token bigram frequency, since that also takes into account both the case of the words and the short and long variants of personal pronouns.

In Table 4 I compiled frequency lists of (token) bigrams consisting of the ten most frequent bigrams of *MIS* + MONOSYLLABIC PRONOUN in newer fiction texts from the years 2013–2017 (based on Raudvere & Uiboaed (2018)), and in spontaneous speech and in dialects (based on the data used in the present study). These pronouns are the same ones that were included in the monosyllabic pronoun value of the following word length variable in the earlier analysis: the short personal pronouns *ma* 'I', *sa* 'you', *ta* 's/he', *me* 'we', *te* 'you (pl)', *nad* 'they', the demonstrative pronouns *see* 'it' and *need* 'these' (including dialectal variants of *too*, *taa* 'it' and *nood*, *naad* 'these') and their monosyllabic case forms.

Table 4. Frequency lists of token bigrams of *MIS* + SHORT (PERSONAL) PRONOUN in fiction texts, spontaneous speech and dialects. The partitive *mida* is in bold text for visualisation purposes.

FICTIONAL TEXTS	SPONTANEOUS SPEECH	DIALECTS
mida *ta*	*mis see*	*mis ta*
mida *ma*	*mis sa*	*mis see*
mida *sa*	*mis ma*	*mis sa*
mis sa	*mis ta*	*mis ma*
mis ta	**mida** *ma*	*mis nad*
mis see	**mida** *me*	*mis need*
mida *see*	**mida** *sa*	*mis me*
mis ma	**mida** *ta*	*mis mul*
mida *nad*	*mis me*	*mis too*
mida *me*	*mis nad*	*mis tal*

Table 4 shows very clear tendencies. In the fictional texts, *mida* is most common in this construction, as the nominative *mis* is more frequent than *mida* only when it is preceding *see* 'it'. In spontaneous speech, the most frequent constructions already involve the nominative *mis*, but the partitive *mida* is still more frequent when it is preceding *me* 'we'. In dialects, however, the ten most frequent constructions all involve the nominative *mis* (the first construction involving a partitive – *mida ma* 'what I' – would be 19th on the frequency list).

These lists obviously illustrate the use of *mis* and *mida* in any position, not only as partial objects, but one can assume that the percentage of the pronouns in partial object position is roughly the same in all the registers, so the conclusions made here based on these frequency lists can most likely also be generalised only for pronouns in partial object position. Moreover, these lists are supported by the distribution of nominative and partitive pronouns in partial object position in the study's datasets (see Table 2).

Why is the range of variation so different in archaic dialects and contemporary speech? I propose that the reason is language standardisation. Although this alternation has not been officially standardised and the exception concerning the object case for the pronoun *mis* has been included in the Estonian grammars at least since 1993, the description of it has not always been completely accurate: both Erelt et al. (1993: 53) and Erelt, Erelt and Ross (2007: 383) state that the variation of *mis* and *mida* in partial object position can only occur in questions, and only in Metslang (2017: 237) has that particular restriction on clause type for this variation been lifted. As this variation definitely exists in all the different clause types in the dialect texts, recorded long before the 1990s, it raises the question whether this restriction in the grammars could have still worked as a subtle way of standardising the phenomenon, since teachers, language planners, editors, etc. could have only relied on official sources in their work. It could be another explanation for why it is precisely questions that include the most nominative pronouns in partial object position.

Whatever the case, it is clear that using the partitive *mida* in partial object position follows the general rules of DOM, while using the nominative *mis* does not. Therefore, it is only logical that fictional texts, which are typically carefully edited and proof-read, have the most frequent use of the partitive *mida*. Spontaneous speech, which is, at times,

somewhat informal in the dataset, is still a language much affected by the standardisation of the literary language, and thus has an equal use of *mis* and *mida* in partial object position. Lastly, the speakers from the dialect corpus speak a language based solely on oral tradition and, moreover, have received very little education during a time when language planning had barely begun in Estonia. As a result, their speech was virtually not influenced by standardisation, which can explain their very prevalent use of the nominative *mis*.

5. Conclusions

In this paper I examined the object case variation of the pronoun *mis* 'what' in the expected partial object position. While nouns in partial object position are characteristically in partitive, the pronoun *mis* is an exception to this rule: the partitive *mida* can be replaced with the nominative *mis* (which is typically the case used for total objects). This variation occurs both in spoken and written Estonian as well as in Estonian dialects. The aim of this paper was to determine which variables affect this object case variation most significantly, how the variation differs between contemporary speech and archaic dialects, and what might have possibly caused or influenced this variation. The analysis was done based on the data from the Phonetic Corpus of Estonian Spontaneous Speech and the Corpus of Estonian Dialects.

While the two corpora differ in their range of variation of *mis* and *mida* in partial object position (in the spontaneous speech data, the division of *mis* and *mida* is more or less equal, but in the dialects, *mida* occurs in less than 10% of the sentences), there are several morphosyntactic and geographical variables that contribute to this variation in both sets of language data. My analyses showed that verb type, clause type, length of the following word, and dialect are the most significant variables affecting this particular variation: the use of the partitive *mida* is more likely in relative clauses, in the case of a partitive verb and in the easternmost dialects, while its use was less likely when followed by a monosyllabic pronoun and in (rhetorical) questions. In addition, differences between individual speakers in the datasets also proved to be relevant.

In this study, the variation of *mis* and *mida* in partial object position has been examined mostly from a morphosyntactic perspective. Yet, it is likely that some prosodic variables, such as intonation and sentence stress, may also contribute to this variation. As discussed earlier, question words used in questions and indirect questions should be stressed and could therefore be longer, i.e., in partitive, while the relative pronouns used in relative clauses should be unstressed, and therefore shorter and in nominative. This is, however, the opposite of what was found in this study, where questions formed the clause type most likely to include a nominative *mis*. Consequently, it is clear that the prosody of *mis*- and *mida*-initial sentences and their variation needs further and more detailed analysis.

As for the reasons behind this variation, I proposed that there could be (at least) three possible motivations for the use of *mis* instead of *mida* in partial object position: language contact, high usage frequency and language standardisation. After further examination, it can be concluded that all three may have had an influence on the variation between *mis* and *mida*. While it is difficult to determine whether any late local contacts have been responsible for the increase of the nominative *mis*, the lack of contacts in the eastern dialects (which have retained the use of the partitive *mida*) could explain why the Coastal, Northeastern and Seto dialects differ from the other dialects. Frequent constructions consisting of *mis* and a monosyllabic pronoun showed a higher percentage of nominative than low-frequency constructions, indicating that high frequency could be the motivation behind the choice to use the shorter nominative in partial object position. Finally, the distinct differences in the range of variation in contemporary speech and archaic dialects could be explained by the effects of language standardisation, which has surely had an influence on the current speech in Estonia, but less so on the dialects coming primarily from an oral tradition.

Acknowledgements

Abbreviations

1, 2, 3 – first, second, third person, ABL – ablative, ADE – adessive, ALL – allative, CED – Corpus of Estonian Dialects, CLI – clitic, CMP – comparative, CNG – connegative, COM – comitative, COND – conditional voice, DISTR – distributive, DOM – differential object marking, ELA – elative, GEN – genitive, IMP – imperative, INE – inessive, INF – infinitive, IPS – impersonal voice, NOM – nominative, PCESS – Phonetic Corpus of Estonian Spontaneous Speech, PL – plural, PRT – partitive, PST – past tense, PTCP – participle, SG – singular

References

Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory* 21(3). 435–483. https://doi.org/10.1023/A:1024109008573.

Alvre, Paul. 1986. Läänemeresoome pronoomenite tüpoloogiat. *Fenno-Ugristica* 13. 5–20.

Alvre, Paul. 1987. Pronoomenite morfoloogiat. Pronoomen *mis*. *Emakeele Seltsi Aastaraamat* 31(1985). 27–36.

Ariste, Paul. 1939. *Eesti-rootsi laensõnadest*. Tartu: K. Mattiesen.

Asu, Eva Liina, Pärtel Lippus, Karl Pajusalu & Pire Teras. 2016. *Eesti keele hääldus* (Eesti Keele Varamu II). Tartu: Tartu Ülikooli Kirjastus.

Baayen, R. Harald, Douglas J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

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. https://doi.org/10.1007/s11185-013-9118-6.

Basile, Rodolfo & Ilmari Ivaska 2021. *Löytyä*-verbin konstruktioiden yhteydessä esiintyvä subjektin sijanvaihtelu. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 12(1). 11–39. https://doi.org/10.12697/jeful.2021.12.1.01.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01.

Berry, Roger. 2018. *English grammar: a resource book for students*. London/New York: Routledge.

Breiman, Leo. 2001. Random forests. *Machine Learning* 45(1). 5–32. https://doi.org/10.1023/A:1010933404324.

Bybee, Joan. 1985. *Morphology: A study of the relation between meaning and form* (Typological Studies in Language 9). Amsterdam/Philadelphia: John Benjamins Publishing.

Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10(5). 425–455. https://doi.org/10.1080/01690969508407111.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Bybee, Joan & Joanne Scheibman. 1999. The effect of usage on degrees of constituency: the reduction of *don't* in English. *Linguistics* 37(4). 575–596. https://doi.org/10.1093/acprof:oso/9780195301571.003.0014.

Bybee, Joan & Dan Slobin. 1982. Rules and schemas in the development and use of the English past tense. *Language* 58(2). 265–289. https://doi.org/10.2307/414099.

Diessel, Holger. 2007. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology* 25(2). 108–127. https://doi.org/10.1016/j.newideapsych.2007.02.002.

Diessel, Holger. 2017. Usage-based linguistics. In Mark Aronoff (ed.), *Oxford Research Encyclopedia of Linguistics*. New York: Oxford University Press.

Durrell, Martin, Katrin Kohl, Gudrun Loftus & Claudia Kaiser. 2013. *Essential German grammar*. London/New York: Routledge.

Erelt, Mati. 2017. Lauseliikmed. Öeldis. In Mati Erelt & Helle Metslang (eds.), *Eesti keele süntaks* (Eesti keele varamu III), 93–239. Tartu: Tartu Ülikooli kirjastus.

Erelt, Mati, Tiiu Erelt & Kristiina Ross. 2007. *Eesti keele käsiraamat*, 3., täiendatud trükk. Tallinn: Eesti Keele Sihtasutus.

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 Keele ja Kirjanduse Instituut.

Gregory, Michelle L., William D. Raymond, Alan Bell, Eric Fosler-Lussier & Daniel Jurafsky. 1999. The effects of collocational strength and contextual predictability in lexical production. In Sabrina J. Billings, John P. Boyle & Aaron M. Griffith (eds.), *Proceedings of the Chicago Linguistic Society Meeting* 35(2), 151–166. Chicago : Chicago Linguistic Society.

Hennoste, Tiit, Andriela Rääbis & Kirsi Laanesoo. 2017. Polar questions, social actions and epistemic stance. *STUF – Language Typology and Universals* 70(3). 139–161. https://doi.org/10.1515/stuf-2017-0023.

Hint, Helen, Piia Taremaa, Maria Reile & Renate Pajusalu 2021. Demonstratiivpronoomenid ja -adverbid määratlejatena. Miks me oleme *siin ilmas*, *selles olukorras*? *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 12(1). 79–111. https://doi.org/10.12697/jeful.2021.12.1.03.

Hooper, Joan B. 1976. Word frequency in lexical diffusion and the source of morphophonological change. In William Christie (ed.), *Current progress in historical linguistics*, 96–105. Amsterdam: North Holland.

Hothorn, Torsten, Kurt Hornik & Achim Zeileis. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15(3). 651–674. https://doi.org/10.1198/106186006X133933.

Ilie, Cornelia. 2015. Questions and questioning. In Karen Tracy, Cornelia Ilie & Todd Sandel (eds.), *The International Encyclopedia of Language and Social Interaction*, 1257–1271. Boston: John Wiley & Sons.

Juhkam, Evi. 1983. Rootsipärane objekt eesti murretes. *Keel ja Kirjandus* 3. 122–125.

Juhkam, Evi & Aldi Sepp. 2000. *Läänemurde tekstid* (Eesti murded VIII). Tallinn: Eesti Keele Instituut.

Jurafsky, Daniel, Alan Bell, Michelle Gregory & William D. Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. *Typological Studies in Language* 45. 229–254. https://doi.org/10.1075/tsl.45.13jur.

Kaalep, Heiki-Jaan & Kadri Muischnek. 2002. *Eesti kirjakeele sagedussõnastik*. Tartu: Tartu Ülikooli Kirjastus.

Keem, Hella & Inge Käsi. 2002. *Võru murde tekstid* (Eesti murded VI). Tallinn: Eesti Keele Instituut.

Kettunen, Lauri. 1943. *Vepsän murteiden lauseopillinen tutkimus* (Suomalais-Ugrilaisen Seuran Toimituksia 86). Helsinki: Suomalais-Ugrilainen Seura.

Klaas, Birute. 1999. Dependence of the object case on the semantics of the verb in Estonian, Finnish, and Lithuanian. In Mati Erelt (ed.), *Estonian: Typological studies III* (Publications of the Department of Estonian of the University of Tartu 11), 47–83. Tartu: Bookmill.

Klavan, Jane. 2021. The alternation between exterior locative cases and postpositions in Estonian web texts. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 12(1). 153–188. https://doi.org/10.12697/jeful.2021.12.1.05

Klavan, Jane, Maarja-Liisa Pilvik & Kristel Uiboaed. 2015. The use of multivariate statistical classification models for predicting constructional choice in spoken, nonstandard varieties of Estonian. *SKY Journal of Linguistics* 28. 187–224.

Kont, Karl. 1963. *Käändsõnaline objekt läänemeresoome keeltes*. Tallinn: Eesti NSV Teaduste Akadeemia.

Koptjevskaja-Tamm, Maria & Bernhard Wälchli. 2001. The Circum-Baltic languages. An areal-typological approach. In Östen Dahl & Maria Koptjevskaja-Tamm (eds.), *Circum-Baltic languages*, *Vol 2: Grammar and typology*, 615–750. Amsterdam/Philadelphia: John Benjamins.

Kortmann, Bernd. 2010. Areal variation in syntax. In Peter Auer & Jürgen Erich Schmidt (eds.), *Language and space: An international handbook of linguistic variation*, *Vol 1: Theories and methods*, 837–864. Berlin/New York: Walter de Gruyter.

Krug, Manfred. 1998. String frequency: A cognitive motivating factor in coalescence, language processing, and linguistic change. *Journal of English Linguistics* 26(4). 286–320. https://doi.org/10.1177/007542429802600402.

Krug, Manfred. 2003. Frequency as a determinant of grammatical variation and change. In Günther Rohdenburg & Britta Mondorf (eds.), *Determinants of grammatical variation in English*, 7–67. Berlin: Mouton de Gruyter.

Laanesoo, Kirsi. 2014. Direktiivsed *mis*- ja *mida*-küsilaused suulises suhtluses. *Emakeele Seltsi aastaraamat* 59(2013). 103–126. https://doi.org/10.3176/esa59.05.

Laanesoo, Kirsi. 2018. *Polüfunktsionaalsed küsilaused eesti argivestluses* (Dissertationes linguisticae universitatis Tartuensis 33). Tartu: Tartu Ülikooli kirjastus.

Lindström, Liina. 2006. Infostruktuuri osast eesti keele sõnajärje muutumisel. *Keel ja Kirjandus* 49(11). 875–888.

Lindström, Liina, Mervi Kalmus, Anneliis Klaus, Liisi Bakhoff & Karl Pajusalu. 2009. Ainsuse 1. isikule viitamine eesti murretes. *Emakeele Seltsi aastaraamat* 54(2008). 159–185.

Lindström, Liina. 2013. Eesti murdekorpus. Center of Estonian Language Resources. https://doi.org/10.15155/1-00-0000-0000-0000-00076L. https://metashare.ut.ee/repository/browse/602bfe185a5111e2a6e4005056b40024160ffef10451449ea0779 8e7c30b3081/.

Lindström, Liina, Maarja-Liisa Pilvik & Helen Plado. 2018. Nimetamiskonstruktsioonid eesti murretes: murdeerinevused või suuline süntaks? *Mäetagused* 70. 91–126. https://doi.org/10.7592/MT2018.70.lindstrom_pilvik_plado.

Lindström, Liina, Maarja-Liisa Pilvik, Mirjam Ruutma & Kristel Uiboaed. 2015. Mineviku liitaegade kasutusest eesti murretes keelekontaktide valguses. *Võro Instituudi toimõndusõq* 29. 39–70.

Lindström, Liina, Maarja-Liisa Pilvik, Mirjam Ruutma & Kristel Uiboaed. 2019. On the use of perfect and pluperfect in Estonian dialects. In Sofia Björklöf & Santra Jantunen (eds.), *Multilingual Finnic – Language contact and change* (Uralica Helsingiensia 14). 155–193. https://doi.org/10.33341/uh.85035.

Lindström, Liina & Kristel Uiboaed. 2017. Syntactic variation in 'need'-constructions in Estonian dialects. *Nordic Journal of Linguistics* 40(3). 313–349. https://doi.org/10.1017/S0332586517000191.

Lindström, Liina, Kristel Uiboaed & Virve-Anneli Vihman. 2014. Varieerumine *tarvis-/vaja*-konstruktsioonides keelekontaktide valguses. *Keel ja Kirjandus* 8–9. 609–630.

Lippus, Pärtel. 2019. Eesti keele spontaanse kõne foneetiline korpus v.1.0.5. Center of Estonian Language Resources. https://doi.org/10.15155/1-00-0000-0000-0000-001A3L. https://metashare.ut.ee/repository/browse/eb031a04998c11e-9a6e4005056b40024d356c0e6b4e4d10ba516a0afb8d67e0.

Metslang, Helle. 2017. Lauseliikmed. Sihitis. In Mati Erelt & Helle Metslang (eds.), *Eesti keele süntaks* (Eesti keele varamu III), 258–277. Tartu: Tartu Ülikooli kirjastus.

Metslang, Helle & Külli Habicht. 2019. Sihitise sirgumine läbi 19. sajandi kirjakeele. *Oma Keel* 38(1). 16–26.

Must, Mari. 1987. *Kirderannikumurre: häälikuline ja grammatiline ülevaade*. Tallinn: Valgus.

Nurme, Maili. 2012. Impersonaali objekt. *Keel ja Kirjandus* 55(5). 335–345.

Ogren, David. 2015. Differential object marking in Estonian: Prototypes, variation, and construction-specificity. *SKY Journal of Linguistics* 28. 277–312.

Ogura, Mieko. 1993. The development of periphrastic *do* in English: A case of lexical diffusion in syntax. *Diachronica* 10(1). 51–85. https://doi.org/10.1075/dia.10.1.04ogu.

Pajupuu, Hille. 1990. Sugulaskeelte intonatsioonist: eesti ja soome keele võrdlus. *Arvutuslingvistika sektori aastaraamat 1988.* 145–159.

Pajusalu, Karl, Tiit Hennoste, Ellen Niit, Peeter Päll & Jüri Viikberg. 2009. *Eesti murded ja kohanimed*, 2., täiendatud trükk. Tallinn: Eesti Keele Sihtasutus.

Pajusalu, Renate. 2006. Multiple motivations for meaning of an elative *wh*-construction in Estonian. *Trames* 10(4). 324–340.

Pook, Hanna. 2019. The pronoun *kes* 'who' and its referent's animacy in Estonian dialects. *SKY Journal of Linguistics* 32. 105–144.

Pool, Raili. 1999. About the use of different forms of the first and second person singular personal pronouns in Estonian cases. In Mati Erelt (ed.), *Estonian: Typological studies III* (Publications of the Department of Estonian of the University of Tartu 11), 158–184. Tartu: Bookmill.

R Core Team. 2018. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rätsep, Huno. 1978. *Eesti keele lihtlausete tüübid*. Tallinn: Valgus.

Raudvere, Uku & Kristel Uiboaed. 2018. *Uuema eesti ilukirjanduse mitmikute loendid*. https://datadoi.ee/handle/33/41.

Reile, Maria. 2015. Space and demonstratives: An experiment with Estonian exophoric demonstratives. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 6(2). 137–165. https://doi.org/10.12697/jeful.2015.6.2.06.

Ruiter, Jan P. de. 2012. Introduction: questions are what they do. In Jan P. de Ruiter (ed.), *Questions: Formal, functional and interactional perspectives*, 1–7. Cambridge: Cambridge University Press.

Rumm, Andra. 2019. *Avatud küsimused ja nende vastused eesti suulises argivestluses*. (Dissertationes linguisticae Universitatis Tartuensis 36). Tartu: Tartu Ülikooli kirjastus.

Ruutma, Mirjam, Aki-Juhani Kyröläinen, Maarja-Liisa Pilvik & Kristel Uiboaed. 2016. Ambipositsioonide morfosüntaktilise varieerumise kirjeldusi kvantitatiivsete profiilide abil. *Keel ja Kirjandus* 2. 92–113.

Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin & Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9(1). 1–11. https://doi.org/10.1186/1471-2105-9-307.

Strobl, Carolin, James Malley & Gerhard Tutz. 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14(4). 323–348. https://doi.org/10.1037/a0016973.

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. https://doi.org/10.1017/S0954394512000129.

Tamm, Anne. 2003. Estonian transitive verb classes, object case, and progressive. *Nordlyd* 31(4). 639–653. https://doi.org/10.7557/12.30.

Tamm, Anne & Natalia Vaiss. 2019. Setting the boundaries: Partitive verbs in Estonian verb classifications. *Eesti Rakenduslingvistika Ühingu aastaraamat* 15. 159–181. https://doi.org/10.5128/ERYa15.09.

Taremaa, Piia. 2017. *Attention meets language: A corpus study on the expression of motion in Estonian* (Dissertationes linguisticae Universitatis Tartuensis 29). Tartu: University of Tartu Press.

Tauli, Valter. 1968. *Keelekorralduse alused*. Stockholm: Eesti Raamat.

Timberlake, Alan. 2004. *A reference grammar of Russian*. New York: Cambridge University Press.

Uiboaed, Kristel. 2013. *Verbiühendid eesti murretes* (Dissertationes philologiae estoni-
cae Universitatis Tartuensis 34). Tartu: Tartu Ülikooli kirjastus.

Uiboaed, Kristel, Cornelius Hasselblatt, Liina Lindström, Kadri Muischnek & John
Nerbonne. 2013. Variation of verbal constructions in Estonian dialects. *Literary and
Linguistic Computing* 28(1). 42–62. https://doi.org/10.1093/llc/fqs053.

Van de Velde, Hans & Roeland van Hout. 1998. Dangerous aggregations. a case study
of Dutch (n) deletion. *Papers in Sociolinguistics*. 137–147.

Viitso, Tiit-Rein & Valts Ernštreits. 2012. Līvõkīel-ēstikīel-leţkīel sõnārõntõz. Liivi-
eesti-läti sõnaraamat. Lībiešu-igauņu-latviešu vārdnīca. Tartu, Rīga: Tartu Ülikool,
Latviešu valodas aģentūra.

Winkler, Eberhard & Karl Pajusalu. 2018. *Salis-Livisch II. Grammatik und Wörterver-
zeichnis. Mit einem Anhang zu den salis-livischen Sprichwörtern.* Wiesbaden:
Harrassowitz Verlag.

Witzlack-Makarevich, Alena & Ilja A. Seržant. 2018. Differential argument marking:
Patterns of variation. In Ilja A. Seržant & Alena Witzlack-Makarevich (eds.), *Dia-
chrony of differential argument marking*. 1–40. Berlin: Language Science Press.
https://doi.org/10.5281/zenodo.1228243.

**Kokkuvõte. Hanna Pook: Pronoomeni *mis* käände varieerumine objekti
positsioonis spontaanses eesti keeles ja eesti murretes.** Eesti keeles eris-
tatakse täis- ja osasihitist mitmete semantiliste ja süntaktilise tunnuste põhjal;
täissihitis on nominatiivis või genitiivis, osasihitis partitiivis. Relatiiv-interro-
gatiivpronoomeni *mis* puhul võib aga oodatud osasihitise positsioonis asen-
dada partitiivi *mida* nominatiiviga *mis*. Selle artikli eesmärk on välja selgitada,
millised tunnused mõjutavad oluliselt pronoomeni *mis* objekti käände varieeru-
mist, kuidas see varieerumine erineb vanemates kohamurretes ja tänapäeva-
ses spontaanses kõnes ning mis on selle varieerumise võimalikud põhjused ja
mõjurid. Analüüs põhineb eesti keele spontaanse kõne foneetilisel korpusel
ja eesti murrete korpusel. Tulemused näitavad, et *mis* ja *mida* varieerumist
osaobjekti positsioonis mõjutavad kõige enam verbitüüp, lausetüüp, järgneva
sõna silpide arv ja murre. On tõenäoline, et pronoomeni *mis* käände varieeru-
mine on korraga olnud mõjutatud mitmest tegurist, peamiselt keelekontaktidest
(või kontaktivähesusest teatud piirkondades), pronoomeni *mis* suurest kasutus-
sagedusest ja keele standardiseerimisest.

Märksõnad: eristav objekti markeerimine, relatiiv-interrogatiivpronoomenid,
süntaks, varieerumine, eesti murded, suuline kõne