The pronoun *kes* 'who' and its referent's animacy in Estonian dialects

Hanna Pook
University of Tartu

Abstract

In the Estonian language the interrogative-relative pronouns *kes* 'who' and *mis* 'what' are distinguished by whether they are used to refer to animate or inanimate entities, respectively. However, in certain Estonian dialects the pronoun *kes* can be used to refer to inanimate entities as well. The aim of this paper is to quantitatively analyse this non-standard use of *kes* based on the data in the Corpus of Estonian Dialects and to determine which dialects typically use *kes* to refer to inanimate entities and which variables (construction type, case marking, number, relative clause position and referential distance) contribute most to this phenomenon. The results show that there are significant differences between the dialects – the pronoun *kes* is mostly used to refer to inanimate entities in Northern Estonia (most frequently in the Eastern dialect, but also in the Mid, Coastal and Western dialects), but this phenomenon is rare in the southern dialects. This paper argues that out of all the variables, the choice to use *kes* to refer to inanimates is most associated with the case of the pronoun, being very frequent in elative and comitative. It also appears that *kes* is used to refer to inanimate entities mostly in postnominal relative clauses and when the word expressing the referent is in plural.

Keywords: dialect syntax, Estonian dialects, interrogative-relative pronouns, animacy

1   Introduction

In many languages there seems to be a strong connection between the animacy of a referent and the choice of ways to refer to it (Dahl & Fraurud 1996: 56). This is also true for the interrogative-relative pronouns *kes* 'who' and *mis* 'what' in Standard Estonian: *kes* is usually used to refer to animate entities and the pronoun *mis* is used to refer to inanimate entities.

However, in some Estonian dialects the pronoun *kes* can be used to refer to inanimate entities (examples 1–2)[1] and *mis* can be used to refer to animate entities.

(1)   Eastern dialect (Avinurme)[2]

*sell   aeal   eij   õld   jällä   noh   seda   **saage   kellega***
that:ADE time:ADE not be:PST:PTCL again PART this:PRT saw:PRT who:COM

puid   leigatta
tree:PL:PRT   cut:INF

'At that time there was no saw to cut trees with.'

(2)   Insular dialect (Käina)

narid   *old*   *kaa* ***kelle***   *sees magadi*
bunk:PL   be:PST:PTCL too who:GEN in   sleep:IPS:PST

'[There] were bunkbeds, too, that were slept in.'

There has been little written about this possibility in previous research. Kask & Palmeos (1985: 59) mention that in some areas of Estonia the relative pronoun *kes* can be used to talk about inanimate entities. This phenomenon is also noted in a few descriptions of certain Estonian subdialects (see Niklus 1957: 139; Pung 1968: 134; Juhkam 2012: 468). However, these previous works only mention the existence of this phenomenon, but do not elaborate on how or why *kes* is used in this way.

The aim of this article is to give a thorough overview of this non-standard use of *kes* for referring to inanimate entities in the Estonian dialects based on the Corpus of Estonian Dialects.[3] The research questions are the following:

1. Which Estonian dialects and subdialects typically refer to inanimate entities with *kes*?

2. Do construction type, the case of the pronoun and the number of the referent affect whether inanimate entities are referred to with *kes*, and if so, how?

---

[1] These and all the following examples are from the Corpus of Estonian Dialects. A brief description of the corpus is given in § 2.1.

[2] Each example is preceded by the dialect and subdialect (in parentheses). The description and map of Estonian dialects and subdialects is given in § 2.1.

[3] This article is largely based on the author's master's thesis *Pronoomeni* kes *kasutusest eesti murretes* ("The use of the pronoun *kes* 'who' in Estonian dialects", Pook 2018), with little corrections and an extended analysis.

3. What is the difference in the use of the pronoun *kes* between pre- and postnominal relative clauses?

4. Is the referential distance between the pronoun *kes* and the referred to entity in the relative clause connected to the animacy of the referent?

This paper is structured as follows: in § 1.1 I give an overview of how animacy is conveyed in languages and in § 1.2 I explain the principles guiding the choice between *kes* and *mis* in Estonian. In § 2 I describe the Estonian dialects, the data found in this study and the methods used. The following analysis of the data is separated into two parts. § 3 describes the association between the analysed variables and the animacy of the referent to which the pronoun *kes* refers, presents the results of this analysis and answers research questions 1 and 2. In § 4 I make a subset of the data containing only those dialect areas that typically use *kes* to refer to inanimate entities (based on the results of § 3) and only headed relative clauses, and analyse those separately. Therefore, § 4 answers research questions 3 and 4. § 5 contains the conclusions and discussion.

1.1   Animacy

Animacy in a language is hardly ever a binary variable like in biology, but regarded more as a continuum that extends from human through animal to inanimate (Yamamoto 1999: 1). The first modern description of this continuum, called the (universal) animacy hierarchy, was presented by Silverstein (1976), but the most common representation of this hierarchy, found in Dixon (1979: 85), is as follows:

> 1st, 2nd person pronoun > 3rd person pronoun > proper names > human common noun > non-human animate common noun > inanimate common noun

Since then, this hierarchy has been used and adapted by countless linguists to describe different construction types in different languages. Some languages and/or constructions use less fine distinctions, e. g. just human versus non-human or animate versus inanimate, while others can have intermediate categories between the levels of the common hierarchy, e. g. humans > higher animals > nonhuman animates or have 1st and 2nd person as separate categories (Dixon 1979: 85; Comrie 1989: 185; Croft 1990: 113).

The categories that are higher in the hierarchy are often grammatically distinguished from those that are lower, are treated as more central to clause structure and are more likely to act as an agent in events (Whaley 1996: 172; Kittilä et al. 2011: 6). Still, this hierarchy is definitely not an ordering of discrete categories, but rather a continuum ranging from most to least animate (Croft 1990: 113).

However, animacy is not the only important parameter reflected in this hierarchy – in fact, there are several others. One of them is empathy: hierarchy assigns primacy to nominals with which the speaker is familiar or has empathy towards (Whaley 1996: 172–173). Certain authors prefer to avoid the term *animacy* altogether and speak rather of *empathy hierarchies*, since there is technically no difference in literal animacy between the 1st person and a proper noun or a human common noun (Kittilä et al. 2011: 6; see also Langacker 1991: 306–307). The empathy hierarchy according to Langacker is the following:

speaker > hearer > human > animal > physical object > abstract entity

Another relevant parameter to the animacy hierarchy is the individuation scale: a stronger sense of animacy is attributed to entities that are seen as individuals than to those that are part of an indeterminate mass (Comrie 1989: 189). The most important distinction in individuality is between singularity and plurality, but there are several examples of 'masses' that lie on the boundary between animacy and inanimacy, e. g. human organisations, geographic entities or communities, which are usually treated as inanimate, but still show certain animate-like features (Yamamoto 1999: 131, 138–139).

All in all, determining whether something is animate or inanimate is not always clear. Not even all animate beings are equally animate to human cognition, since animacy has a gradience from centrally animate beings to peripherally animate beings (Yamamoto 1999: 14). According to Fowler (1977: 16–17), animate beings are those that are capable of initiating actions and change, either through conscious or unconscious drive (whereas inanimate beings lack this capability to cause the world to change), and of movement (which is why plants are typically perceived as inanimate). In this study I follow Fowler's distinction, categorising all humans (and human collectives) as well as animals as animate. Therefore this paper mainly focuses on the animacy hierarchy instead of the empathy hierarchy, since it is generally possible to categorise the entities in the dataset by their biological animacy. A more detailed description of the data coding can be found in § 2.2.

1.2   *kes* and *mis* in Estonian

According to Erelt (2017: 743), the choice between *kes* 'who' and *mis* 'what' in Standard Estonian is determined by the animacy hierarchy:

> humans > higher animals > lower animals[4] > inanimates

and the individuation scale:

> an individual > a distributive group > a collective

With these hierarchies, the probability of using *kes* decreases and the probability of using *mis* increases from left to right. The relative pronoun *kes* is always (in Standard Estonian) used for a single human referent and the relative pronoun *mis* is always used to refer to an inanimate entity. In the case of multiple referents, *kes* is preferred for a distributive group and *mis* is preferred for a collective group. The choice of the pronoun also depends on whether the collective or institution is in the role of agent or patient – in the first case, *kes* is preferred, while in the second case, *mis* is preferred. It can be generalised that the more active the role of the referred to entity, the more likely the use of the pronoun *kes* is (Erelt 2014: 743).

Animate non-human entities can be referred to using both *kes* and *mis*. The pronoun *mis* can even be used to refer to a single animal: with higher animals, the use of *mis* is justified if the animal is not the agent in the sentence, for lower animals the choice of the pronoun is freer and it is always possible to use the pronoun *mis* (Erelt et al. 2007: 561).

Examples that run contrary to these principles have not only been found in Estonian dialects, but are also apparent in old literary Estonian: while in those texts *kes* usually refers to animate entities, there are certain relative clauses where the relative pronoun *kes* refers to inanimate referents, such as *usk* 'belief', *puu* 'tree', *kiri* 'letter', etc. (Kõpp 2001: 41–42). According to Erelt (1996: 11) the choice between the pronouns *kes* and *mis* has not always functioned according to contemporary norms, even in the written texts of the 19th and beginning of the 20th century.

In addition, Wiedemann (2011: 470) notes in his Estonian grammar written in 1875 that in Mid-Estonia only the interrogative pronoun *kes* (referring to animates) was carefully kept apart from the interrogative pronoun *mis* (referring to inanimates), while using these pronouns in a relative clause, a

---

[4] Erelt does not specify what he considers to be a higher or a lower animal.

similar distinction was often not made (however, he also remarks that in some dialects the distinction was not made for interrogative pronouns either).

It is possible that a clear distinction of interrogative pronouns, which was not followed as precisely with relative pronouns, is related to the grammaticalisation of the interrogative marker. In Estonian (and in most Finno-Ugric languages) interrogative sentences are an older construction than relative clauses, which means that the interrogative pronouns are older than the relative pronouns (Alvre 1987: 23). According to Heine & Kuteva (2006: 209), interrogative markers have four possible stages of grammaticalisation, from being just an interrogative marker to introducing headed relative clauses:

1. Only an interrogative marker (*Who came?*).

2. Interrogative marker is extended to introduce indefinite complements or adverbial clauses (*I don't know **who** came.*).

3. In addition to indefinite clauses, the interrogative marker can introduce definite complement or adverbial clauses, which can also be interpreted as headless relative causes (*You also know **who** came.*).

4. In addition to definite non-headed relative clauses, the interrogative marker can introduce headed relative clauses (*Do you know the woman **who** came?*).

Because the interrogative marker in Estonian (and in most Balto-Finnic languages) has gone through all of these grammaticalisation stages, the words for both interrogative and relative pronouns are the same (Erelt 2017: 683). It has been noted about many languages that going down the stages decreases the distinctions in gender, animacy, number and case and the marker assumes a more general (grammatical) meaning (Heine & Kuteva 2006: 227).

Based on that, we can deduce that in Estonian the pronouns *kes* 'who' and *mis* 'what' are the most systematically distinguished by animacy in independent interrogative sentences and the least in relative clauses. Therefore, relative clauses could be the construction type where *kes* most probably refers to inanimate entities, because in such constructions the meaning of the interrogative marker is probably the most abstract and unclear. In this paper I attempt to find evidence to support this theory.[5]

---

[5] It should be noted that *kes* and *mis* are not the only pronouns that have lost some of their

2   Data and methods

2.1   Estonian dialects, Corpus of Estonian Dialects and data

The analysed data has been collected from the Corpus of Estonian Dialects (CED),[6] which contains authentic spoken texts from all Estonian dialects. There are several different ways that the dialects can be divided and grouped, but in this corpus they have been divided into ten traditional dialect areas: the Mid, Western, Insular and Eastern dialects (i. e. the North-Estonian dialect group), the Coastal and Northeastern dialects (i. e. the Northeastern-Coastal dialect group), and the Tartu, Võru, Mulgi and Seto dialects (i. e. the South-Estonian dialect group) (Lindström 2015). These dialect areas can be seen in Figure 1.

The most significant contrast exists between the northern (North-Estonian and Northeastern-Coastal dialect groups) and southern dialects, with the greatest differences being found in their phonology, morphology and lexis. According to K. Pajusalu et al. (2009: 73), the oldest distinctive features of these two dialect regions can be traced back thousands of years, when they were the differences between Balto-Finnic tribal languages. Still, most of the distinct features developed later due to language contacts, administrative borders, etc. South-Estonian dialects are at times even considered to be a separate language. Recent studies in dialect syntax have, however, found that for certain constructions the main isogloss line can be drawn rather between the western and eastern dialects (see e. g. Uiboaed 2013; Uiboaed et al. 2013; Lindström et al. 2014; 2015; 2018).

All the dialects in the corpus have been divided into subdialects (the borders of which are based on former administrative units – parishes). The subdialects represented in the corpus data that is used in the present article are shown in Figure 2. It is clear that the representation of subdialects differs between dialects – e. g. the data contains all of the subdialects of the Northeastern and Võru dialect, but only about half of the subdialects are

---

distinction in animacy in spoken Estonian. In Standard Estonian, the personal pronoun *tema*/*ta* 's/he' is used to refer to animate entities and the demonstrative pronoun *see* 'it' is used to refer to inanimates. However, in spoken Estonian, in the case of a very familiar entity or in some idiomatic expressions, *ta* (the short form of the pronoun) is often used to refer to inanimates. Similarly, the demonstrative *see* is used to refer to animate entities either in a presentational clause, when there are two different animate referents involved or for emphasis in the case of a relatively new referent (R. Pajusalu 2005: 110, 132–133).

[6] http://www.murre.ut.ee/estonian-dialect-corpus/ (accessed 2020-04-16).
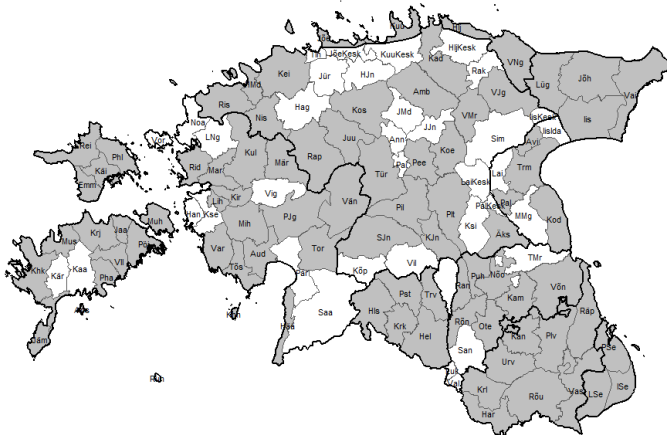
Figure 1. The dialect division of the CED

Figure 2. Subdialects in the analysed data: grey for subdialects that are represented in the data, white for the subdialects not represented in the data. Thicker lines represent the borders between dialects.

Table 1. The number of the informants, total tokens and keyword *kes* in the dialects

dialect	informants	total tokens	keyword *kes*
Western	62	251,031	741
Mid	70	246,167	680
Insular	53	202,325	621
Võru	33	111,503	389
Coastal	24	97,152	329
Eastern	20	48,353	205
Seto	23	68,414	195
Tartu	21	80,343	191
Mulgi	16	63,759	187
Northeastern	15	60,037	171
Σ	337	1,229,084	3,709

represented in the Mid dialect. It should be noted, however, that the Mid dialect area is one of the biggest dialect areas, so there are still more texts from that dialect (and other bigger dialects that are not represented in this data in their entirety) than from, for example, the Northeastern or Võru dialect.

The CED includes dialect recordings, phonetically transcribed texts, dialect texts in simplified transcriptions, morphologically annotated texts and a separate database for information about the informants and recordings. This analysis uses (manually) morphologically annotated dialect texts, from which all the uses of the pronoun *kes* (based on the lemma 'kes') have been collected into the data, in total 3,709 keywords. Every keyword is accompanied by the preceding and following context (up to 20 words), case marking (already annotated in the corpus texts) and information about the informant (dialect, subdialect, village, age, date of birth, gender and place of birth).

The frequency of the keyword *kes* in different dialects is presented in Table 1. The table also includes the number of informants and the total number of tokens in the texts from each dialect. As expected, the number of informants and the length of the texts mostly correspond with the frequency of the keyword. The number of the informants and total tokens varies so much across the dialects mainly because some dialects areas are much bigger than others and some dialects have had more available informants.

Out of the 337 informants 236 were women and 101 were men. They were

born between 1864 and 1922 and ranged in age from 49 to 100 years old. The texts were recorded mostly in the 1960s and 1970s, but the earliest recordings date to the year 1938 (Lindström 2015). This means that the language analysed in this paper is antiquated (although the dialects are more or less still in use in the rural areas of Estonia, especially in Southern Estonia and in the insular region).

2.2   Data coding

The data attained from the corpus contains, as mentioned above, keywords, context, case and information about the informant. In addition, I have manually added certain variables. I did this in two stages. In the first stage I added the referent's animacy, construction type and number of the word expressing the referent to every sentence in the data. I also labelled whether relative clauses were headed or non-headed.

Animacy's variable has two levels: animate and inanimate. As previously mentioned, I have followed Fowler's (1977: 16–17) distinction in determining animacy: all humans (and human collectives) as well as animals are categorised as animate, while everything else is categorised as inanimate. Due to the nature of the analysed texts in the CED (which typically include topics like the informant's personal life, their lifestyle, past events or working methods), the marking of the referred to entities as animate or inanimate was relatively straightforward and there were hardly any borderline cases of animacy.

The sentences were divided into seven construction types based on which types most frequently appeared in the data. A detailed description of all the construction types is presented in § 3.2.

The number of the word expressing the referent is either singular or plural. (It should be noted that this is not the number of the pronoun, since the pronoun *kes* does not have plural forms and is only used in singular.)

In the second stage, in the second part of the analysis, I added two variables only to headed relative clauses: distinguishing whether the relative clause is pre- or postnominal as well as measuring the distance in words between the pronoun *kes* and the word expressing the referent. The specifics of those variables are explained thoroughly in § 4.1 and § 4.2. All the variables, their levels and abbreviations are presented in Table 2.

The CED consists of spoken texts, which means a significant number of sentences in the dataset are incomplete or contain repetition of the pronoun

Table 2. The variables in the dataset and their levels. If applicable, the abbreviations of the levels used in subsequent graphs are given in parentheses.

variable	levels
dialect	Eastern (E), Western (W), Mid (MID), Insular (I), Coastal (C), Northeastern (NE), Mulgi (MUL), Tartu (T), Võru (V), Seto (S)
animacy	animate (anim), inanimate (inanim)
construction type (TYPE)	relative clause (rel), relative clause without a main clause (rel_main), question (q), indirect question (indir_q), rhetorical question (rhet_q), listing construction (list), other
case marking (CASE)	nominative (nom), genitive (gen), partitive (prt), elative (ela), allative (all), adessive (ade), translative (trl), comitative (com)
number (NUM)	singular (S), plural (P)
relative clause position (REL_POS)	prenominal (pre), postnominal (post)
referential distance in words (DISTANCE)	0, 1, 2, 3, 4, 5, 6, 7+

kes. Since it was not possible to analyse unfinished sentences and including repetitions did not add anything to the analysis, these sentences were excluded from the dataset.

In the end, the final dataset consisted of 3,324 keywords *kes* with context.

2.3   Methods

This study is a corpus-based dialect analysis, which has many advantages over classic atlas-based or dictionary-based approaches, some of which I will expand on here. The corpus is based on natural (spoken) language use, so the data from the corpus is authentic and reliable (whereas atlas data can sometimes be biased, since it has mostly been collected by questionnaires and therefore can only represent conscious language use). Corpus-based dialectology takes into account the variation and the frequency of different variants, so instead of being able to say that a certain feature exists in one dialect and does not exist in another, we can calculate the exact frequencies

and differences between dialects. The corpora also include the immediate context around the feature, which gives an opportunity to analyse semantic, pragmatic and textual characteristics of the feature (Lindström & Pilvik 2018: 649).

There are also some restrictions to the data when using corpus-based approaches (see Szmrecsanyi & Anderwald 2018: 302), but overall the possibility to systematically look at the features and the extent of their usage gives us a much wider picture of the variation than previous dialect atlas data could.

To determine which variables most influence the decision to use the pronoun *kes* to refer to inanimate referents, I have analysed the data in two ways: firstly I have looked at the results quantitatively, inspecting the frequencies of the values and (if possible) finding significant correlations between them using the Chi-Squared test. Secondly, I have applied the conditional inference tree method. This was chosen due to its suitability for analysing the data, which contains mostly nominal variables that are often not acceptable as inputs using other statistical methods. Unlike parametric tests, this method does not require any distributional assumptions to be met (which this dataset does not). Conditional inference trees (and random forests) have also been successfully used in other studies about Estonian dialect syntax (see Klavan et al. 2015; Ruutma et al. 2016; Lindström & Uiboaed 2017; Lindström et al. 2018, etc.).

The conditional inference tree method is based on binary recursive partitioning, wherein each stage the algorithm tests whether any of the independent variables are 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 (Levshina 2015: 291).

All the calculations for the results were performed using the program R (R Core Team 2018). The conditional inference trees were computed using the function `ctree()` in the package `party` (Hothorn et al. 2006).

3   Impact of the analysed variables

In this section of the paper, I analyse different variables to find out how they affect the decision to use the pronoun *kes* to refer to inanimate entities. The analysis has been divided into two parts. In this section (§ 3) I examine how the animacy of the referent is associated with dialects and subdialects, different construction types, the case of the pronoun and the number of the referent.

In the following section (§ 4) I extract a subset from the data that only consists of headed relative clauses and those dialect areas that (according to the results of § 3) use *kes* frequently to refer to inanimate entities, to see whether pre- and postnominal relative clauses or the referential distance between the pronoun and the word used to express the referred to entity have any differences in regards to the usage of the pronoun *kes*.

## 3.1   Dialects and subdialects

In order to find the differences in the use of the pronoun *kes* between the dialects, I compared the dialects and subdialects represented in the data in terms of how frequently inanimate entities were referred to using the pronoun *kes*. The hypothesis was that referral to inanimate entities using *kes* is typical in the North-Estonian dialect group, particularly in the Mid and Western dialects, but does not occur in the South-Estonian dialect group. This hypothesis was based on a previous automated study of the same data, which provided adequate results to establish this hypothesis. However, that study made it clear that an automated analysis of spoken text cannot give conclusive results. Therefore, for this study all the material was analysed manually.

The frequency of animate and inanimate referents that the pronoun *kes* is referring to is separated by dialect in Table 3. The percentages of animate and inanimate referents are presented in Figure 3. Using the pronoun *kes* to refer to inanimate entities is most frequent in the Eastern (14.4% of all the pronoun *kes* occurrences were used to refer to an inanimate entity, 28 pronouns out of 194), Mid (14.2%, 87 out of 613), Coastal (13.8%, 41 out of 297) and Western dialects (13%, 88 out of 676). The feature is quite rare in the Tartu (5.8%, 10 out of 173), Northeastern (5.7%, 9 out of 159) and Võru dialects (2.4%, 8 out of 336), and it does not occur at all in the Seto dialect.

This confirms the hypothesis that referring to inanimate entities with *kes* is more frequent in the North-Estonian dialect group, but the percentages are clearly not homogenous within all the dialect groups. The Insular

Table 3. The raw frequency of animate and inanimate referents in dialects

dialect	animate	inanimate	Σ
Mid	526	87	613
Western	588	88	676
Insular	515	36	551
Eastern	166	28	194
Coastal	256	41	297
Northeastern	150	9	159
Mulgi	145	16	161
Tartu	163	10	173
Võru	328	8	336
Seto	164	0	164
Σ	3,001	323	3,324
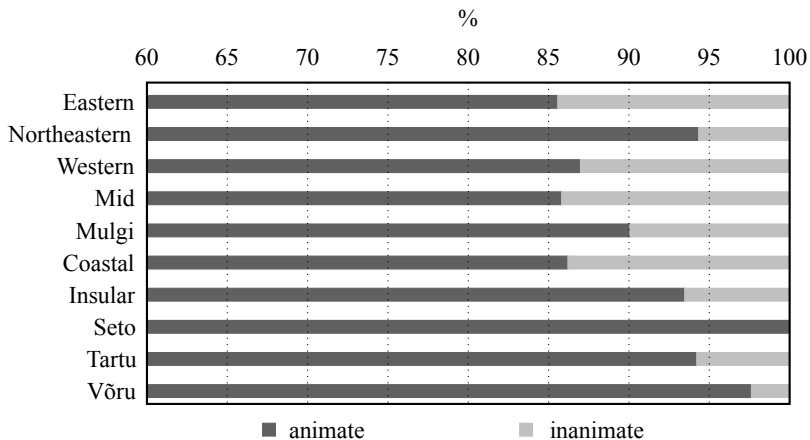
Figure 3. The percentage of animate and inanimate entities referred to by the pronoun *kes* in different Estonian dialects

dialect differs the most from the rest of the North-Eastern dialect group. Moreover, the Coastal and Northeastern dialects, which both belong to the Northeastern-Coastal dialect group (and are even categorised as one Coastal dialect in the Estonian Dialect Dictionary[7]), differ significantly in regard to this particular phenomenon. Therefore, it might not be possible to completely categorise this use of *kes* according to the traditional dialect groups.

Nevertheless, we can say that the South-Estonian dialect group stands out the most from the other groups. It can thus be concluded that in the South-Estonian dialects the pronoun *kes* is used differently from other dialect groups and the referred to entity is almost always animate. The examples in the data where *kes* refers to inanimate entities in those dialects are more exceptions than a typical use of the pronoun. There were significantly more inanimate referents in the Mulgi dialect than in other South-Estonian dialects, but those were mostly due to the idiolect of one informant. However, other studies in dialect syntax have found that the Mulgi dialect cannot always be grouped together with the other southern dialects, but rather behaves more similarly to the northern dialects (see Lindström et al. 2015; Ruutma 2016), which might also be the reason for the Mulgi dialect's distinctiveness.

In the following map of all the subdialects and the percentage of the referenced inanimate entities (Figure 4), it can be seen that the use of the pronoun *kes* to refer to inanimate entities is most frequent in Northern and Central Estonia, but is considerably less frequent in Southern and Eastern Estonia and on the island of Saaremaa. However, based on the subdialects represented in the data, there does not seem to be a specific dialect area even in Northern Estonia where this use of *kes* is more prevalent than elsewhere. Instead, there are single subdialects (or groups of subdialects) all over Northern Estonia that tend to use this feature more, bordered by subdialects that rarely or never use *kes* to refer to inanimate entities.

Looking at the subdialects more closely, it appears that using *kes* to refer to inanimate entities is mostly widespread all over the Eastern, Coastal, Western and Mid dialects, with very few exceptions. The subdialects in which the most inanimate entities were referred to using *kes* were Avinurme and Palamuse from the Eastern dialect (20% [6 pronouns out of 30] and 19.1% [16 out of 84] of all the referenced entities were inanimate, respectively), Juuru and Koeru from the Mid dialect (46.9% [15 out of 32] and 24% [12 out of 50], respectively), Haljala and Jõelähtme from the Coastal dialect (26.2% [11 out

---

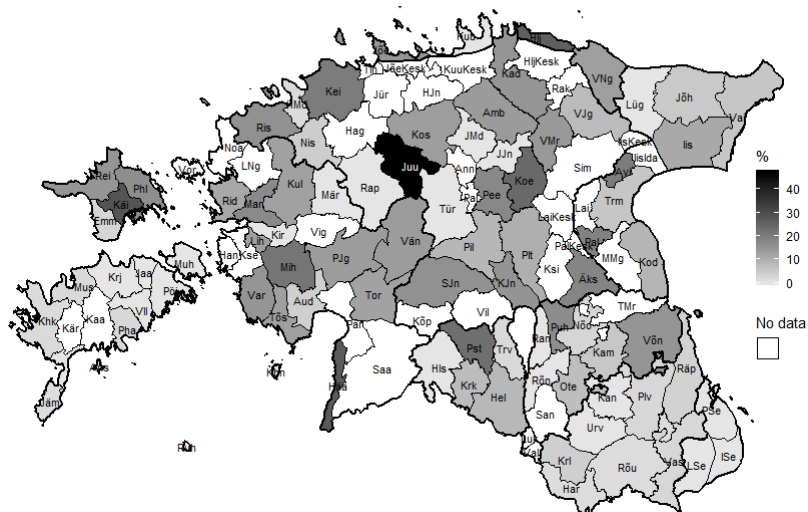[7] http://www.eki.ee/dict/ems/ (accessed 2020-04-16).

Figure 4. The percentage of the pronoun *kes* used to refer to inanimate entities in the represented subdialects. The thicker lines represent the borders between dialects.

of 42] and 15.1% [18 out of 119], respectively) and Häädemeeste and Mihkli from the Western dialect (27.3% [6 out of 22] and 21.4% [6 out of 28], respectively).

Even though the average percentage of inanimate referents is low in the Insular dialect (only 6.5%, 36 pronouns out of 551) and most subdialects do not refer to an inanimate entity with *kes* at all, there are three subdialects that considerably raise the overall percentage of the dialect: Käina (28.6%, 18 out of 63), Pühalepa (16.4%, 10 out of 61) and Reigi (15%, 3 out of 20). All these subdialects are spoken on the island of Hiiumaa. Therefore, while the rest of the Insular dialect uses *kes* more like the South-Estonian dialect group, Hiiumaa's subdialects clearly have the option of using *kes* to refer to inanimate entities.

The reason for this is not clear – it is possible that using *kes* to refer both to animate and inanimate entities in those subdialects has been influenced by contacts with the Swedish language, which uses one main relative pronoun

som to refer to both animate and inanimate entities (Tuldava 1993: 144–147). Inanimate entities in Hiiumaa's subdialects are mainly referred to by *kes* in relative clauses, so this could be one explanation for the frequent use of *kes*.

However, language contact is a wide topic and its effect is difficult to measure without a larger study. In a previous study of mine (Pook 2018) one of the research questions was whether some of the contact or cognate languages have influenced the use of the pronoun *kes* in Estonian dialects, but it was not possible to come to any clear conclusions about that. Therefore, the possible effect of language contact is excluded from the scope of this paper.

It should be noted, nevertheless, that language contact has been said to affect and even accelerate grammaticalisation (Heine & Kuteva 2005: 14), which, as described in § 1.2, is one likely reason behind the phenomena studied in this paper. For example, language contact induced grammaticalisation has played an important role in the development of articles in Eastern European languages, including the article-like demonstrative pronoun *see* 'it' and numeral *üks* 'one' in Estonian language (Hint et al. 2017: 68–69; see also R. Pajusalu 1997; 2009; Heine & Kuteva 2006). So while language contact is not discussed in this paper, upon further reading one must keep in mind that it may not only be a possible reason behind the pronoun *kes* being used to refer to inanimates, but it may also be a hidden factor behind the grammaticalisation of the pronoun.

3.2   Construction type

In this section, I look at how the factor of construction type is related to the informants' choice to use the pronoun *kes* for referring to inanimate entities. I divided all the sentences into seven types, basing the division on types most frequently represented in the data. These included very common construction types like relative clauses (3), interrogative sentences (5), rhetorical questions (6) and indirect questions (7), but also some that are unique to the data or to the spoken language, such as relative clauses lacking a main clause that it is supposed to be modifying (4; these kinds of constructions either had the function of introducing a new topic or lacked a main clause simply because the speaker's train of thought changed mid-sentence), repeated referencing while listing (8; repeatedly using *kes* to refer to a part of a group) and other constructions that did not belong to any of the previous types.[8]

---

[8] The entities referred to by *kes* in these examples are all animate, since the purpose of them is to illustrate construction types and not the studied phenomena.

(3)   Northeastern dialect (Jõhvi)

*igal   maeal   kaks ärga   **kes** olid   künni   jaust*
every:ADE house:ADE two ox:PRT who be:PST:3PL plow:GEN for

'Every house [had] two oxen that were for plowing.'

(4)   Mid dialect (Põltsamaa)

kell   *ol*   *maa nigu mull   ol   maa mull*
who:ADE have:PST:3SG land like me:ADE have:PST:3SG land me:ADE

ess   ole   tuuga   aigugi   tettä   ess
not:PST have:CNG that:COM time:PRT:CLI do:INF not:PST

'[S/he] who had land like I had land, I did not even have time to do [anything] with that.'

(5)   Western dialect (Martna)

kelle   *müts oli   siis see selle   Lauri   vana müts või*
who:GEN hat   be:PST:3SG then this that:GEN Lauri:GEN old hat   or

'Whose hat was this, that Lauri's old hat?'

(6)   Insular dialect (Muhu)

*nee   oo   meelest juba   kadun   **kiss** teab   kus*
that:PL be:3SG mind:ELA already disappear:PST:PTCL who know:3SG where

naad läin   oo
they go:PST:PTCL be:3SG

'Those are gone from memory already, who knows where they have gone.'

(7)   Võru dialect (Põlva)

*eit tiijäq   **kess** tu pesä   sõss ol   maaha ajanu*
not know:CNG who that nest:GEN then be:PST:3SG down push:PST:PTCL

'[I] don't know who had pushed that nest onto the ground.'

(8)   Seto dialect (Eastern-Seto)

kess *kudi   kinnast   **kes** kudi   sukka*
who knit:PST:3SG glove:PRT who knit:PST:3SG stocking:PRT

'Who knitted a glove, who knitted stockings.'

The hypothesis was that referring to inanimates with *kes* is most common in relative clauses, since in the process of the interrogative markers'

grammaticalisation (from being just an interrogative marker to being able to introduce headed relative clauses) they could have lost their distinction in animacy, as described by Heine & Kuteva (2006: 209, 227). The detailed reasoning behind this has already been explained in § 1.2.

We can see in Table 4 that the majority of the data consists of relative clauses: 71.3 % of the constructions are relative clauses.

Relative clauses also include the highest percentage of inanimate entities referred to by *kes* (12.4%), confirming my hypothesis. Pronouns in other construction types were rarely (or not at all) used to refer to inanimate entities. The percentage of the pronoun *kes* being used to refer to inanimate entities by construction type is depicted in Figure 5.

The data contained a significant number of non-headed relative clauses (i. e. relative clauses lacking a head noun they were modifying; example 3 headed, example 9 non-headed). Despite that omitting the head is grammatically incorrect in Standard Estonian, it is quite frequent in spoken language. There were 341 non-headed relative clauses in the data, of which only 10 contained an inanimate entity (10–11).

(9) Mulgi dialect (Helme)

kes tegi   parembat   rohkemp tüüd   sai   ka rohkemp
who do:PST:3SG good:CMP:PRT more   work:PRT get:PST:3SG also more
palkka
pay:PRT

'[S/he] who did better [and] more work also got more pay.'

(10) Mid dialect (Suure-Jaani)

omale *jäi*   *kaa* **keda**   *sa* *jäll*   *tuleva*
self:ALL remain:PST:3SG too who:PRT you again come:PTCL:GEN
kevadi   *maha külisid*
spring:GEN down sow:PST:2SG

'[Some] was left for you as well that you sowed again next spring.'

(11) Western dialect (Tõstamaa)

see oli   **kellega** *linad*   *pehmes peksetti*
it   be:PST:3SG who:COM flax:PL soft:TRL hit:IPS:PST

'This was [the one] that was used to soften flax.'

Out of all the relative clauses 14.4% were non-headed (341 out of 2,371),

Table 4. The raw frequencies of animate and inanimate referents in different construction types

construction type	animate	inanimate	Σ
relative clause	2,078	293	2,371
relative clause without a main clause	221	5	226
interrogative sentence	76	5	81
rhetoric question	335	13	335
indirect question	160	6	166
listing construction	60	0	60
other	71	1	72
Σ	3,001	323	3,324
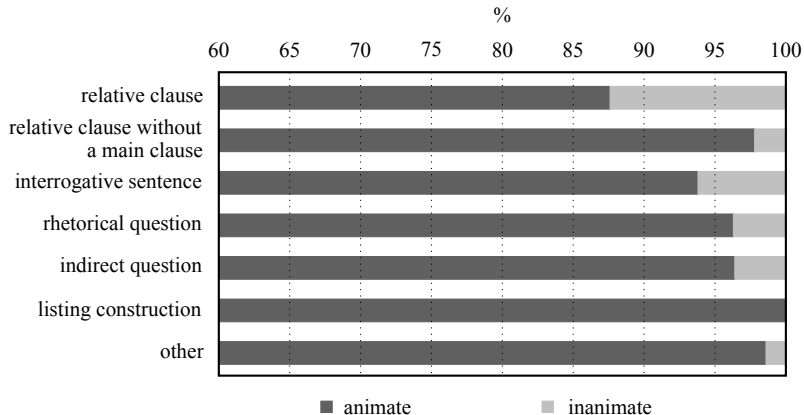
Figure 5. The percentage of animate and inanimate entities referred to by the pronoun *kes* in different construction types

but out of the relative clauses that modified an inanimate entity, only 3.4% were headless (10 out of 293). It is possible that when the pronoun *kes* is used to refer to an inanimate entity, omitting the head of the relative clause from the sentence might cause some misunderstandings and that is why in those cases informants almost always felt the need to clearly state the head in the sentence. Overall, from this analysis we can conclude that out of all construction types, *kes* is most typically used to refer to inanimate entities in headed relative clauses.

3.3   Case marking

To find out how certain grammatical cases are associated with the use of *kes* in regards to referring to inanimate entities, I analysed the cases of all the pronouns in the data.[9] Eight of the Estonian language's fourteen cases were represented in the data: the nominative, genitive, partitive, elative, allative, adessive, translative and comitative cases. Pronouns in these cases were used to refer to both singular and plural entities (the pronoun *kes* does not itself have plural forms and is always used in the singular), but in this section I have not made a distinction in number. (The relation between the number of the word expressing the referent and the informants' decision to use *kes* to refer to inanimate referents is addressed in § 3.4.) The frequency of the cases of the pronoun *kes* in the data separated by animacy are presented in Table 5.

As expected, most of the pronouns in the data were in the nominative case (2,398 pronouns out of 3,324, 72.1%), but using that case to refer to inanimate entities was quite rare: out of all the pronouns in nominative, only 3.2% (76 pronouns) referred to an inanimate entity. The percentage of inanimate entities in all of the cases is represented in Figure 6. It can be seen that every case in the data (except translative) was used to refer both to animate and inanimate entities, meaning that using *kes* to refer to inanimate entities is not limited to certain cases, but is used across the entire paradigm.

Clearly two of the cases stand out from the others: elative, which had 90.6% of all the pronouns referring to an inanimate entity (29 pronouns out of 32), and comitative, which had 78.9% of all the pronouns referring to an inanimate entity (164 out of 208). Note that for every other variable analysed previously, the percentage of animate referents has been dominant, far larger than the percentage of inanimate referents. With the elative and comitative

---

[9] For a thorough overview of the Estonian case system, see e. g. Blevins (2008).

Table 5. The frequency of animate and inanimate referents in different cases

case	animate	inanimate	Σ
nominative	2,322	76	2,398
genitive	88	32	120
partitive	57	6	63
elative	3	29	32
allative	33	1	34
adessive	453	15	468
translative	1	0	1
comitative	44	164	208
Σ	3,001	323	3,324
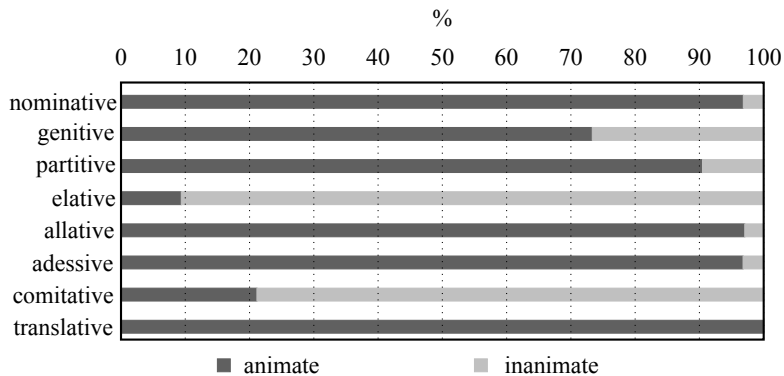
Figure 6. The percentage of animate and inanimate entities referred to using the pronoun *kes* in different cases

cases, however, the percentage of the referred to inanimate entities greatly exceeds the number of referred to animate entities, which means that these two cases (or rather the relations they represent) could possibly affect the choice of pronoun the most out of all the variables analysed thus far.

Since elative and comitative differed so much from the other cases, I proceeded to examine their semantic functions in the dataset more thoroughly.

The semantic function of the comitative case is to represent a companion, an instrument, time, manner, condition, etc. (Erelt et al. 2007: 253). In this dataset the pronoun *kes* in the comitative case was only used to refer to inanimate entities when the referent's function in the sentence was to be an instrument (12–13). The comitative case's function of being an instrument has developed from the companion's function through the metaphor AN INSTRUMENT IS A COMPANION (Metslang et al. 2017: 157; see also Stolz et al. 2009). As the instrument may be said to be more grammaticalised than the companion, since the latter usually implies a human participant, whereas the former does not (Heine et al. 1991: 157), a word in the function of an instrument can therefore expectedly be referred to with the animate pronoun in the comitative case.

(12)   Mid dialect (Keila)

*sealt   pealt sai   vett   võtta   ja **kellega** pesu*
from_there on   get:PST:3SG water:PRT take:INF yes who:COM laundry:PRT

pesta
wash:INF

'From there one could take water to wash laundry with.'

(13)   Western dialect (Kullamaa)

*niukse   vankri   tegin   **kellega**   kellega   saap*
that_kind:GEN wagon:GEN make:PST:1SG who:COM   who:COM get:3SG

kiva   põllalt   ära   vädada
stone:PL:PRT field:ABL away carry:INF

'[I] made this kind of a wagon that can be used to carry stones off the field.'

The elative case has many functions in a sentence: it can signify a source, starting time, starting state, material, stimulus, etc. (Erelt et al. 2007: 248–249). In Standard Estonian, the prototypical function of elative is a source or a starting time. However, in this dataset elative was mainly used in the function of material (14). Since there were only three sentences in which elative was

in the function of a source (and even these do not contain a location in the traditional sense; see 15), it can be concluded that *kes* is preferred when referring to entities that are concrete and not locations or abstract times.

(14)   Mid dialect (Juuru)

see   pidi   nii pieenikke ja   ilus   olema see   lõng
it   have_to:PST:3SG so fine   and beautiful be:SUP that thread

kellest   *kootti*
who:ELA   knit:IPS:PST

'It had to be so fine and beautiful, the thread that was used to knit.'

(15)   Mid dialect (Kose)

ta old   ikke üks nõu   kõrvad kõrvad   külges ***kellest***
it be:PST:PTCL PART one container handle:PL handle:PL attached who:ELA

sa   tõstsid   teda
you lift:PST:2SG he:PRT

'It was a container with attached handles that you lifted it with.'

In fact, when all the inanimate entities in the data referred to by *kes* (altogether 323) were categorised as abstract or concrete, only 13 pronouns referred to an abstract inanimate entity (16–17), while 95% of all the pronouns referred to a concrete inanimate referent. This shows that the use of the pronoun *kes* has not expanded into the pronoun *mis*'s function of referring to abstract things.

(16)   Western dialect (Kullamaa)

põlluarimisest *või või* ***kellest***   *see akkab*
farming:ELA   or   or   who:ELA it   start:3SG

'From farming or what it starts from.'

(17)   Mid dialect (Põltsamaa)

luubainaea *oli*   ***kiss*** *sellas   käis*
nightmare   be:PST:3SG who back:INE walk:PST:3SG

'[It] was the nightmare that haunted (lit. walked on one's back).'

3.4   Number

In this section I analyse how the number of the referent is associated with the choice to use the pronoun *kes* to refer to inanimates. To do this, I added the

Table 6. The frequency of animate and inanimate referents separated by the number of the referent

number	animate	inanimate	Σ
singular	2,178	199	2,377
plural	823	124	947
Σ	3,001	323	3,324

number of the word expressing the referred to entity to each sentence. It should be noted that when the referent was only implied in the sentence, but no word for it was included (mostly in the case of questions and indirect questions), the number was always marked as singular if there were no clues in the sentence indicating otherwise.

I hypothesised that since the choice between *kes* 'who' and *mis* 'what' is in part determined by individuation scale (an individual > a distributive group > a collective), where the likelihood of using *kes* decreases and the likelihood of using *mis* increases from left to right, there should be a higher percentage of plural referents among inanimate referents than among animate referents, since in Standard Estonian the right side of the scale is usually covered by the pronoun *mis* and while referring to inanimates, *kes* is used in the function of *mis*.

The results of the analysis are presented in Table 6. It can be seen that 27.4% of animate referents are in plural, while among inanimate referents the percentage is much higher: 38.4%. The Chi-Squared test finds a significant association between the animacy and the number of the referent ($\chi^2 = 16.68$, $df = 1$, $p = 0.000044$), although according to Cramér's V ($V = 0.07$) the association between these two variables is not strong at all. Nevertheless, the results support the established hypothesis.

3.5   The interactions of the variables

In this section I conduct a conditional inference tree analysis for my data. The aim of this analysis is to figure out how the variables that I previously analysed separately are associated with the animacy of the referent in relation to each other. This method shows which of the variables influence the choice of using *kes* most significantly and how the variables interact with each other.
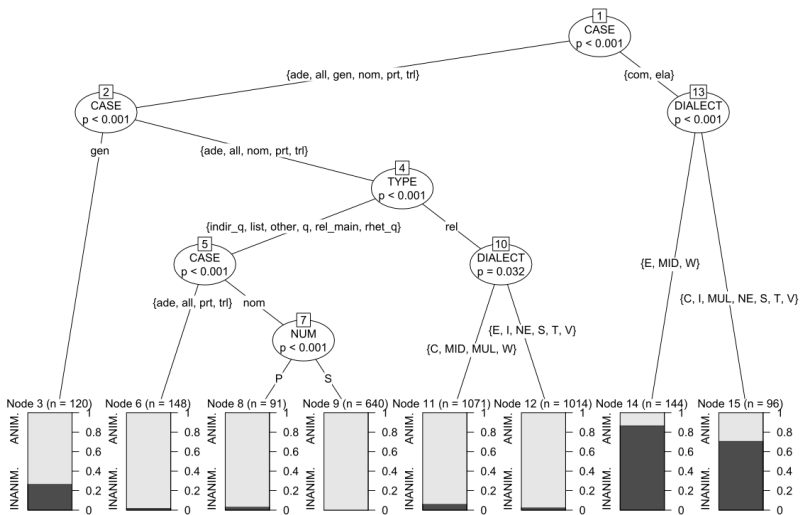
Figure 7. Conditional inference tree for the animacy of the entity that *kes* is referring to

Figure 7 displays the conditional inference tree graph for the animacy of the entity that *kes* is referring to. The variables in this model are the dialect, the construction type, the case and the number of the word expressing the referent. The splits are made in each node by the variable that is most strongly associated with the response: the animacy of the referent. The figure presents all the possible splits significant at the level of 0.05. The bar plots at the bottom show the proportions of animate (light grey) and inanimate (dark grey) referents in each end node, which contain all observations with the given combination of features.

As can be seen in Figure 7, the variable most associated with the animacy of the referent is the case of the pronoun *kes*, which divides the dataset into two based on whether the pronoun is in the elative or comitative case, or some other case. This is in accordance with the preceding analysis which also separated these two cases from the others. Within the group of elative and comitative, a further split is made by dialect: in the Eastern, Mid and Western

dialects, the possibility of an inanimate referent is higher in these two cases (Node 14) than it is in all the other dialects (Node 15), but the percentage of inanimate referents is still relatively high for both of the dialect groups. This is mostly reflective of previous results, except for the Coastal dialect, which typically does use *kes* to refer to inanimate entities in comitative and elative.

The splitting of variables is more complex on the left branch of the tree for the adessive, allative, genitive, nominative, partitive and translative cases. The next significant division is again made by case, which separates genitive from the rest (Node 3), as it is used to refer to inanimates more than the other cases (recall that genitive pronouns referred to the third highest number of inanimate referents).

The sorting of adessive, allative, nominative, partitive and translative on the graph is made by a division of construction types: relative clauses are separated from the rest of the constructions. Relative clauses are divided again by dialect: the Coastal, Mid, Mulgi and Western dialects have a slightly higher percentage of the pronoun *kes* referring to an inanimate entity in relative clauses (Node 11) than the rest of the dialects do (Node 12), although the percentages of inanimate entities are low for both of these groups.

The other construction types (all but the relative clause) are again divided by case, separating nominative from adessive, allative, partitive and translative (Node 6). For the pronouns in nominative the final split is made by the number of the referent: it is possible (although rarely) to refer to an inanimate referent, if the word expressing that referent is in plural (Node 8), but there are no words expressing inanimate referents that are in singular and nominative (Node 9).

The classification accuracy of this tree is 0.95 and the C-index of concordance is 0.89, which both indicate a very good fit.

This analysis shows that the variables included in the method have quite complex interactions with each other. Most of the results correspond with what was discovered before – referring to inanimate entities is most likely to occur in elative and comitative, somewhat likely to occur in genitive, and most prevalent in the Eastern, Mid and Western dialects, although the variable of the dialect is not always significant in affecting the choice to use the pronoun *kes* to refer to inanimate entities. It also showed that for elative, comitative and genitive the construction type is not a significant divider: in those cases, *kes* is used to refer to inanimates in all of the construction types.

Table 7. The number of subdialects, animate and inanimate referents separated by dialect in the new subset

dialect	subdialects	animate	inanimate	Σ
Eastern	3	73	23	96
Mid	15	236	74	310
Western	11	280	74	354
Coastal	3	120	36	156
Insular	3	51	28	79
Σ	35	760	235	995

4 A subset of headed relative clauses

Since the majority of the constructions in the dataset were relative clauses (2,375, 71.4% of all the constructions), I analysed them separately to see whether there are any other variables that affect the use of *kes*. To do that, I made a subset of all the headed relative clauses in the data.

To obtain adequate results about the difference between the use of *kes* in regards to referring to animates and inanimates, I limited this subset to only those dialects that (according to the previous analysis) typically use *kes* to refer to inanimate entities, and of those dialects only the subdialects that had at least 8% of the pronouns referring to inanimate entities. This subset still included sentences with both animate and inanimate entities, but only in those subdialects that had a significant variation in the use of *kes*. The characteristics of this new dataset are described in Table 7.

4.1 Pre- and postnominal relative clauses

Firstly, I intended to find out whether the position of the relative clause in relation to the main clause is connected with the referred to entity's animacy. To that end, I divided all the headed relative clauses according to whether they were preceding (prenominal; 18) or following (postnominal; 19) the main clause (for more about relative clause types, see e. g. De Vries 2001). In the Estonian language the postnominal relative clauses are typically more frequent (Erelt 2004: 407–408), but both options are grammatically correct and both are used in the spoken language.

Table 8. The frequency of prenominal and postnominal relative clauses and referent's animacy separated by dialects

dialect	prenominal		postnominal		Σ
	animate	inanimate	animate	inanimate	
Eastern	14	0	59	23	96
Mid	71	1	165	73	310
Western	69	1	211	73	354
Coastal	16	1	104	35	156
Insular	20	2	31	26	79
Σ	190	5	570	230	995

(18)   Coastal dialect (Jõelähtme)

kes *sis   kõige kangemb   oli   et   **sie** sai   voidu*
who then most   strong:CMP be:PST:3SG that this get:PST:3SG victory:GEN

omale
for_him/herself:ALL

'Who was the strongest, that one won.'

(19)   Western dialect (Martna)

*selle   jaoks on   niisukke **labidas kelle**   peale   paned   jahh*
that:GEN   for   be:3SG this_kind spade   who:GEN   on:ALL put:2SG yes

ja   lükkad   ahju
and push:2SG oven:ILL

'For that there is this kind of a spade that [you] put [the bread] on and put [it] into the oven.'

The results of the distribution of the relative clauses by position and animacy in the data can be seen in Table 8. Just like in Standard Estonian, the postnominal relative clauses are also prevalent in the dialects. However, for relative clauses that modify a word expressing inanimate entity, the postnominal relative clause seems to be almost the only option used. Only five of the inanimate heads precede a relative clause, while 25% of the animate heads precede a relative clause.

One explanation for this could be, once again, the grammaticalisation of the interrogative marker. If the pronoun *kes* is right next to the word expressing

the referred to entity, then its purpose in the sentence is not to add any semantic value, but just to fill the position of the relative marker to form the sentence. Therefore, the closer the pronoun is to the head, the less important its contrast between animate and inanimate is and it is not always necessary to use a pronoun corresponding to the animacy of the referent. However, when the relative pronoun is further from the head, like it is in the case of the prenominal relative clause, it needs to be more semantically connected in order for the listener to understand which entity the pronoun is referring to (see also Heine & Kuteva 2006: 60–61).

4.2   Referential distance

In this section I look at referential distance between the pronoun *kes* and the word expressing the referred to entity in the sentence. This is why this subset of relative clauses included only headed relative clauses: to look at referential distance at all, a sentence has to have both the pronoun and the head noun that the relative clause is modifying. The goal is to see whether this distance has any effect on choice of the pronoun.

The method for implementing referential distance in the analysis of topic continuity was devised by Givón (1983: 11–13), whose method uses this measurement to assess the difficulty of identifying the referent: the shorter the distance is between references to the same participant, the easier it is to identify that participant. Givón's method measures the distance between participants in clauses, one clause being the minimal possible value (in which case the participant is referred to in the previous clause). This referential distance has been found to be significant in other studies about Estonian dialects and spoken Estonian (see e. g. Kalmus 2009; Lindström et al. 2009; Lindström & Vihman 2017).

In this study, however, I measure the referential distance between the pronoun and the head noun in words instead of clauses: for each relative clause the distance in words was counted between the head of the relative clause and the pronoun *kes* (example 19 with the referential distance of zero, example 20 with a referential distance of five). Note that for prenominal relative clauses, the counting started at the end of the relative clause, i. e. the words in the relative clause were not included in the count. The frequencies of the distances are presented in Table 9.

Table 9. The frequency of the distance (in words) between the pronoun *kes* and the head of the relative clause, separated by animacy, including the percentage of the referential distances respectively for animate and inanimate entities.

distance	animate		inanimate		Σ
	n	%	n	%	
0	430	56.6	97	41.3	527
1	156	20.5	64	27.2	220
2	80	10.5	39	16.6	119
3	38	5.0	22	9.4	60
4	23	3.0	5	2.1	28
5	13	1.7	3	1.3	16
6	11	1.5	3	1.3	14
7+	9	1.2	2	0.9	11
Σ	760	100.0	235	100.0	995

(20) Mid dialect (Ambla)

*aga **tädi** suri   tänavu   tall   ära nüid **kiss** teda   ikke*
but   aunt die:PST:3SG this_year s/he:ADE off now who s/he:PRT always

rohkem sis   uolitses   ja   kasvattas
more   then take_care:PST:3SG and raise:PST:3SG

'But his aunt died this year who used to take care of him and raise him.'

Table 9 shows that for both animate and inanimate entities the preferred position of the pronoun is still as close as possible to the head of the relative clause. However, for inanimate entities the distance is distributed more equally and there are more pronouns that are at a one-, two- or three-word distance from the head. Even though the Chi-Squared test confirms that there is a significant association between referential distance and entity's animacy ($\chi^2 = 23.72$, $df = 7$, $p = 0.0013$), the association between these two variables is not strong enough (Cramér's $V = 0.15$) to draw any definite conclusions.

4.3 The interactions of the variables in headed relative clauses

In this section I conduct a conditional inference tree analysis for the relative clause subset, to determine how the variables that could be coded only for
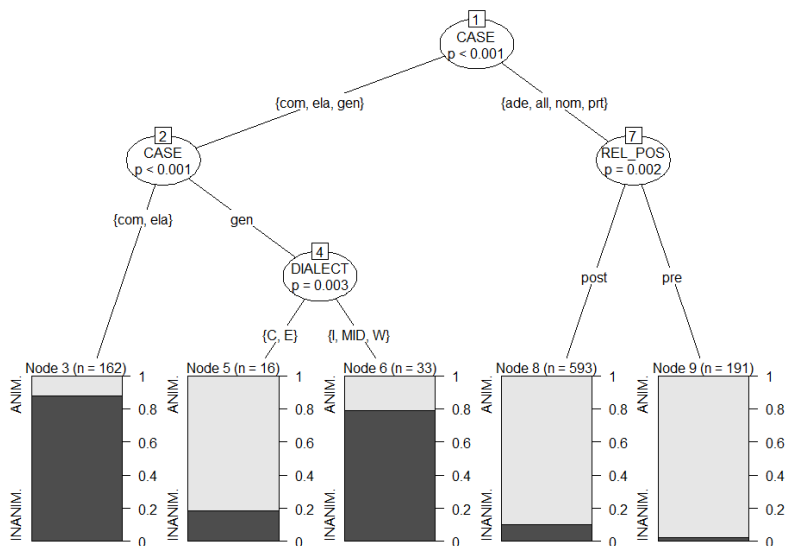
Figure 8. Conditional inference tree for the animacy of the entity that *kes* is referring to in headed relative clauses

headed relative clauses are associated with the animacy of the referent and whether the other variables act similarly to how they did in the conditional inference tree model for the entire dataset.

Figure 8 displays the conditional inference tree graph for the animacy of the entity that *kes* is referring to in headed relative clauses. The variables in this model are the dialect, the case, the number of the word expressing the referent, position of the relative clause relative to the main clause and the referential distance between the pronoun and the head of the relative clause. Again, the splits in each node are made by the variable that is most strongly associated with the response: the animacy of the referent. The bar plots at the bottom show the proportions of animate (light grey) and inanimate (dark grey) referents in each end node, which contain all observations with the given combination of features.

As was the case with the previous conditional inference tree model, the

variable most associated with the animacy of the referent is the case of the pronoun *kes*. The split between the cases is slightly different: elative, genitive and comitative are first grouped together; however, their next split is again made by case and it separates genitive from elative and comitative, which both have a very high percentage of pronouns referring to inanimates (Node 3). The pronouns in genitive are next separated by dialect: *kes* refers to inanimate entities rarely in the Coastal and Eastern dialects (Node 5), but more frequently in the Insular, Mid and Western dialects (Node 6).

For the rest of the cases (adessive, allative, nominative and partitive) the split is made according to the position of the relative clause: the pronoun *kes* is more likely to refer to inanimate entities with a postnominal relative clause (Node 8) than with a prenominal relative clause (Node 9). The percentages for both of these nodes are quite low; nevertheless, it shows that the position of the relative clause only becomes important in those pronoun *kes* cases that are typically not used to refer to inanimates. A closer look at the dataset shows that there are hardly any prenominal relative clauses that have the pronoun *kes* in comitative or genitive (and none in elative), regardless of whether the referred to entity is animate or inanimate. This shows that the use of prenominal relative clauses is not only restricted by the animacy of the referent, but also by the case of the pronoun.

The classification accuracy of this tree is 0.91 and the C-index of concordance is 0.87, which again indicate a very good fit.

As can be seen from the graph, the referential distance between the pronoun and the head of the relative clause (that is, the referred to entity) did not appear to be significant in this model. This confirms my previous conclusion that the distance, while seemingly significantly different for animate and inanimate referents according to the Chi-Squared test, is not actually meaningfully associated with the animacy of the referent or the choice of the pronoun.

The number of the word expressing the referent was also not significant in this conditional tree model, although it proved to be significantly associated with the animacy of the referent in the general analysis of the variable. However, in this subset of the data the animate and inanimate referents are divided very similarly by the number. In fact, for both animate and inanimate referents, 61% of the words expressing the referents were in singular and 39% were in plural, so it is clear that the referent's number is only important in other construction types, but not in headed relative clauses.

5   Discussion and conclusions

In this paper I have examined the way the pronoun *kes* is used in Estonian dialects. While there is a clear distinction between the pronouns *kes* 'who' and *mis* 'what' in Standard Estonian, in Estonian dialects it is not always so: *kes* can sometimes refer to inanimate entities and *mis* to animate entities. The aim of this paper was to determine which Estonian dialects typically use the pronoun *kes* to refer to inanimate entities and which variables most affect this use.

The first part of the analysis showed that inanimate entities were typically referred to with *kes* in the entire North-Estonian dialect group (except on the island of Saaremaa) and in the Coastal dialect. Overall, however, it was not possible to categorise the dialects and subdialects in a way they have been usually divided in previous studies based on other features. Although southern dialects could be excluded from the typical area of use, in the other dialects there was no clear area where *kes* was used to refer to inanimates: there were isolated subdialect groups all over Northern Estonia where the feature was frequently used. Quite a few dialects and subdialects that are otherwise similar (or even overlapping) differed significantly in the case of this feature (e. g. the Coastal and Northeastern dialects). Nonetheless, it is probable that this use of *kes* has been still typically spread all over Northern Estonia, but is just not always represented in the corpus data used.

Out of all construction types, relative clauses included the highest percentage of inanimate entities that were referred to by the pronoun *kes*, while there were hardly any instances of this with other construction types. I speculate this is due to the grammaticalisation of the interrogative marker, which in the position of introducing (headed) relative clauses is in its most grammaticalised form and has therefore lost some of its distinction in animacy.

When referring to inanimates, the case of the pronoun was mostly elative and comitative, at times also genitive, but the use of other cases was rare. The conditional inference tree analysis showed that while the different variables had complex interactions with each other, the choice to use the pronoun *kes* to refer to inanimate referents was mainly associated with the case of the pronoun (or rather the relations those cases represented). There was also a higher percentage of plural words expressing the referents among inanimate referents than among animate referents.

The second part of the analysis, which included only headed relative

clauses, showed that in the case of inanimate entities only postnomimal relative clauses were used, while relative clauses that modified a word expressing an animate entity were both pre- and postnominal. This could, once again, be due to the grammaticalisation of the interrogative marker, whose purpose in postnominal relative clauses (where the pronoun immediately or closely follows the word expressing the referred entity) is not to add any semantic value, but to just fill the position of the relative marker, and therefore its clear distinction in animacy is not at all as relevant as it would be in prenominal relative clauses.

However, the conditional inference tree showed that the position of the relative clause was only significant for those case markings that were typically not used to refer to inanimates. It also appeared that the referential distance between the pronoun and the word expressing the referred to entity was not significantly associated with the choice of the pronoun or the animacy of the referent.

This all shows that although the pronoun *kes* is mainly used to refer to animate entities, there is still significant variation in its use and animacy as a category is not as semantically significant in the dialects as it is in Standard Estonian. It is thanks to corpus-based analysis (which is rather new in the study of dialects) that we are able to examine this (and other) features systematically: prior to this study, there were only scattered notes about the pronoun *kes* being used to refer to inanimate referents, but now it is possible to more clearly define the extent of this feature both in terms of dialect area and sentence context.

In the long term, corpus-based analysis could also help to broaden our understanding of the effects of language contact, which could explain some of the pronoun *kes*'s variation (as several of the contact and cognate languages use interrogative-relative pronouns independently of the referent's animacy). Currently this would be difficult to analyse, due to the fact that the specific features of the local varieties of other languages in Estonia at that time are unknown – we mostly only know the usage of the standard versions of those languages.

Nevertheless it is clear that analysing the counterpart of the pronoun *kes* – the pronoun *mis* 'what' – is necessary. Previous dialect overviews (e. g. Kask & Palmeos 1985: 60; Lonn & Niit 2002: 55, etc.) have claimed that *mis* can also refer to animate entities (as opposed to inanimate entities like in Standard Estonian) and the data of this paper shows that *kes* and *mis* are often used to refer to the same entity by the same informant, either in a self-repair clause

or in consecutive finite clauses. Therefore, further analysis would allow us to get a more comprehensive picture of interrogative-relative pronoun variation based on the referent's animacy.

Acknowledgements

Abbreviations

1, 2, 3	first, second, third person
ABL	ablative
ADE	adessive
ALL	allative
CLI	clitic
CMP	comparative
CNG	connegative
COM	comitative
ELA	elative
GEN	genitive
ILL	illative
INE	inessive
INF	infinitive
IPS	impersonal voice
PART	particle
PL	plural
PRT	partitive
PST	past tense
PTCL	participle
SG	singular
SUP	supine
TRL	translative

References

Alvre, Paul. 1987. *Soome-ugri keelte ajalooline grammatika I: Sõnaliigid: Arvu-ja käändekategooria* [Historical grammar of Finno-Ugric languages I: Parts of speech: The categories of number and case]. 2nd printing. Tartu: Tartu Riiklik Ülikool.

Blevins, James P. 2008. Declension classes in Estonian. *Linguistica Uralica* 44(4). 241–267.

Comrie, Bernard. 1989. *Language universals and linguistic typology: Syntax and morphology*. Oxford: Blackwell.

Croft, William. 1990. *Typology and universals*. Cambridge: Cambridge University Press.

Dahl, Östen & Fraurud, Kari. 1996. Animacy in grammar and discourse. In Fretheim, Thorstein & Gundel, Jeanette K. (eds.), *Reference and referent accessibility*, 47–64. Amsterdam: John Benjamins.

De Vries, Mark. 2001. Patterns of relative clauses. *Linguistics in the Netherlands* 18(1). 231–243.

Dixon, Robert M. W. 1979. Ergativity. *Language* 55(1). 59–138.

Erelt, Mati. 1996. Relative words in Estonian relative clauses. *Tartu Ülikooli eesti keele õppetooli toimetised* 4. 9–23.

— 2004. Märkmeid eesti keele komplekslause kohta [Some notes on the Estonian complex sentence]. *Keel ja Kirjandus* 6/2004. 401–413.

— 2014. *Eesti keele lauseõpetus: Komplekslause* [Estonian syntax: Complex sentence]. Tartu: Tartu Ülikooli eesti keele osakond.

— 2017. Liitlause [Complex sentence]. In Erelt, Mati & Metslang, Helle (eds.), *Eesti keele süntaks* [Syntax of the Estonian language], 647–755. (Eesti Keele Varamu III). Tartu: Tartu Ülikooli Kirjastus.

Erelt, Mati & Erelt, Tiiu & Ross, Kristiina. 2007. *Eesti keele käsiraamat* [Handbook of Estonian]. 3rd, improved edn. Tallinn: Eesti Keele Sihtasutus.

Fowler, Roger. 1977. *Linguistics and novel*. London: Routledge.

Givón, Talmy. 1983. Introduction. In Givón, Talmy (ed.), *Topic continuity in discourse: A quantitative cross-language study*, 5–41. (Typological Studies in Language 3). Amsterdam: John Benjamins.

Heine, Bernd & Claudi, Ulrike & Hünnemeyer, Friederike. 1991. *Grammaticalization: A conceptual framework*. University of Chicago Press.

Heine, Bernd & Kuteva, Tania. 2005. *Language contact and grammatical change*. Cambridge University Press.

— 2006. *The changing languages of Europe*. Oxford: Oxford University Press.

Hint, Helen & Nahkola, Tiina & Pajusalu, Renate. 2017. With or without articles? A comparison of article-like determiners in Estonian and Finnish. *Lähivõrdlusi. Lähivertailuja* 27. 65–106.

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

Juhkam, Evi. 2012. *Harju-Madise murrak* [The subdialect of Harju-Madise]. Ed. by Mari-Liis Kalvik & Helmi Neetar. Tallinn: Eesti Keele Sihtasutus.

Kalmus, Mervi. 2009. *Isikuühildumine eesti murretes* [Person agreement in Estonian dialects]. Tartu: Tartu Ülikool. (Master's thesis).

Kask, Arnold & Palmeos, Paula. 1985. *Eesti keele grammatika 2: Morfoloogia: Arv- ja asesõnad* [Estonian grammar 2: Morphology: Numerals and pronouns]. 3rd printing. Tartu: Tartu Riiklik Ülikool.

Kittilä, Seppo & Västi, Katja & Ylikoski, Jussi. 2011. Introduction to case, animacy and semantic roles. In Kittilä, Seppo & Västi, Katja & Ylikoski, Jussi (eds.), *Case, animacy and semantic roles*, 1–26. (Typological Studies in Language 99). Amsterdam: John Benjamins.

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

Kõpp, Kadri. 2001. *Interrogatiiv-relatiivpronoomenid eesti vanemas kirjakeeles* [Interrogative-relative pronouns in old literary Estonian]. Tartu: Tartu Ülikool. (Bachelor's thesis).

Langacker, Ronald W. 1991. *Foundations of cognitive grammar, vol. 2: Descriptive application*. Stanford: Stanford University Press.

Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.

Lindström, Liina. 2015. *Ülevaade eesti murrete korpusest* [Overview of the Corpus of Estonian Dialects].
(https://www.keel.ut.ee/sites/default/files/www_ut/emk_teejuht2015.pdf)
(Accessed 2020-04-16).

Lindström, Liina & Kalmus, Mervi & Klaus, Anneliis & Bakhoff, Liisi & Pajusalu, Karl. 2009. Ainsuse 1. isikule viitamine eesti murretes [The first person singular reference in Estonian dialects]. *Emakeele Seltsi aastaraamat* 54 (2008). 159–185.

Lindström, Liina & Pilvik, Maarja-Liisa. 2018. Korpuspõhine kvantitatiivne dialektoloogia [Corpus-based quantitative dialectology]. *Keel ja Kirjandus* 8–9/2018. 643–662.

Lindström, Liina & Pilvik, Maarja-Liisa & Plado, Helen. 2018. Nimetamiskonstruktsioonid eesti murretes: Murdeerinevused või suuline süntaks? [Naming constructions in Estonian dialects: Dialectal differences or spoken syntax?] *Mäetagused* 70. 91–126.

Lindström, Liina & Pilvik, Maarja-Liisa & Ruutma, Mirjam & Uiboaed, Kristel. 2015. Mineviku liitaegade kasutusest eesti murretes keelekontaktide valguses [The

use of the compound past tenses in light of language contacts]. *Võro Instituudi toimõtisõq* 29. 39–70.

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

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

Lindström, Liina & Vihman, Virve-Anneli. 2017. Who needs it? Variation in experiencer marking in Estonian 'need'-constructions. *Journal of Linguistics* 53(4). 789–822.

Lonn, Varje & Niit, Ellen. 2002. *Saarte murde tekstid* [Insular dialect texts]. (Eesti Murded VII). Tallinn: Eesti Keele Instituut.

Metslang, Helle & Habicht, Külli & Hennoste, Tiit & Jürine, Anni & Laanesoo, Kirsi & Ogren, David. 2017. Artikkel õnnesoovidega: Komitatiivi funktsioonidest eri aegade ja registrite eesti kirjakeeles [An article with felicitations: The functions of comitative in literary Estonian of different times and registers]. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 8(1). 149–178.

Niklus, Lea. 1957. *Otepää murrak* [The subdialect of Otepää]. Tartu: Tartu Riiklik Ülikool. (Diploma thesis).

Pajusalu, Karl & Hennoste, Tiit & Niit, Ellen & Päll, Peeter & Viikberg, Jüri. 2009. *Eesti murded ja kohanimed* [Estonian dialects and place names]. Ed. by Mari-Liis Kalvik & Tiit Hennoste. 2nd, improved edn. Tallinn: Eesti Keele Sihtasutus.

Pajusalu, Renate. 1997. Is there an article in (spoken) Estonian? In Erelt, Mati (ed.), *Estonian: Typological studies II*, 146–177. Tartu: Tartu Ülikooli Kirjastus.

— 2005. Anaphoric pronouns in spoken Estonian. In Laury, Ritva (ed.), *Minimal reference: The use of pronouns in Finnish and Estonian discourse*, 107–134. Helsinki: Suomalaisen Kirjallisuuden Seura.

— 2009. Pronouns and reference in Estonian. *Sprachtypologie und Universalienforschung* 62(1–2). 122–139.

Pook, Hanna. 2018. *Pronoomeni* kes *kasutusest eesti murretes* [The use of the pronoun *kes* 'who' in Estonian dialects]. Tartu: Tartu Ülikool. (Master's thesis).

Pung, Lea. 1968. *Tartu-Maarja murrak* [The subdialect of Tartu-Maarja]. Tartu: Tartu Riiklik Ülikool. (Diploma thesis).

R Core Team. 2018. *R: A language and environment for statistical computing.* (https://www.R-project.org/) (Accessed 2020-04-16).

Ruutma, Mirjam. 2016. *Kaassõnad eesti murretes* [Adpositions in Estonian dialects]. Tartu: Tartu Ülikool. (Master's thesis).

Ruutma, Mirjam & Kyröläinen, Aki-Juhani & Pilvik, Maarja-Liisa & Uiboaed, Kristel. 2016. Ambipositsioonide morfosüntaktilise varieerumise kirjeldusi

kvantitatiivsete profiilide abil [Descriptions of the morphosyntactic variation of ambipositions by means of quantitative profiles]. *Keel ja Kirjandus* 2/2016. 92–113.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Dixon, Robert M. W. (ed.), *Grammatical categories in Australian languages*, 112–171. Canberra: Australian National University.

Stolz, Thomas & Stroh, Cornelia & Urdze, Aina. 2009. Varieties of comitative. In Malchukov, Andrej L. & Spencer, Andrew (eds.), *The Oxford handbook of case*, 601–608. New York: Oxford University Press.

Szmrecsanyi, Benedikt & Anderwald, Lieselotte. 2018. Corpus-based approaches to dialect study. In Boberg, Charles & Nerbonne, John & Watt, Dominic (eds.), *The handbook of dialectology*, 300–313. Hoboken: John Wiley & Sons.

Tuldava, Juhan. 1993. *Rootsi keele grammatika* [Swedish grammar]. 4th printing. Tartu: Tartu Ülikool.

Uiboaed, Kristel. 2013. *Verbiühendid eesti murretes* [Verbal constructions in Estonian dialects]. Tartu: Tartu Ülikool. (Doctoral dissertation).

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

Whaley, Lindsay J. 1996. *Introduction to typology: The unity and diversity of language*. Thousand Oaks: Sage.

Wiedemann, Ferdinand Johann. 2011. *Eesti keele grammatika* [Estonian grammar]. Ed. by Ellen Niit. Trans. by Heli Laanekask. Tallinn: Eesti Teaduste Akadeemia Emakeele Selts. (Original German edition published in 1875).

Yamamoto, Mutsumi. 1999. *Animacy and reference: A cognitive approach to corpus linguistics*. Amsterdam: John Benjamins.

Contact information:

Hanna Pook
University of Tartu
Institute of Estonian and General Linguistics
Jakobi 2
51014 Tartu
Estonia
e-mail: hanna(dot)pook(at)gmail(dot)com