Nordic Journal of Linguistics (2024), **47**, 192–223
doi:10.1017/S0332586522000221

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

The use of the indefinite pronoun *keegi* 'someone' in Estonian dialects

Hanna Pook and Liina Lindström

Institute of Estonian and General Linguistics, University of Tartu, Jakobi 2, 51005 Tartu, Estonia
Emails for correspondence: hanna.pook@ut.ee and liina.lindstrom@ut.ee

(Received 20 April 2022; revised 29 August 2022; accepted 21 September 2022; first published online 28 October 2022)

Abstract

The Estonian indefinite pronouns *keegi* 'someone' and *miski* 'something' are distinguished by being able to refer to animate or inanimate entities, respectively. However, in certain Estonian dialects, *keegi* is used to refer to inanimate entities as well. The aim of this paper is to describe the functions and use of *keegi* based on the data in the Corpus of Estonian Dialects. We used statistical analyses to determine which dialects typically use *keegi* to refer to inanimate entities and which variables (polarity, function, position in the clause, case marking) contribute most to this variation. The results show that there are significant differences between the dialects: *keegi* is mostly used to refer to inanimate entities in the northern dialects (most frequently in the Western, Mid, and Eastern dialects), but this phenomenon is rare or non-existent in the southern dialects. All of the variables studied contribute to this variation: *keegi* is most likely to refer to an inanimate being when it is in the partitive case, functions as an object, a partitive subject, or a negative polarity item, and is positioned at the end of a negative clause.

Keywords: animacy; dialect syntax; Estonian dialects; indefinite pronouns; negation; spoken language; variation

1. Introduction

Indefinite pronouns, as their name suggests, are pronominal words whose main function is to express indefinite reference (Haspelmath 1997:11), such as *nothing*, *someone*, *anywhere*, etc. in English. In Estonian they typically refer to an undefined or unknown object, phenomenon, or characteristic (Erelt, Erelt & Ross 2007:187).

This study focuses on the use of the indefinite pronoun *keegi* in Estonian dialects, which can have multiple functions depending on the context and polarity of the sentence, and corresponds to the English indefinite pronouns *someone*, *nobody/no one*, *anybody*, etc., as illustrated by the following examples from Standard Estonian.

(1) **Keegi** tuli.
someone *come.PST.3SG*[1]
'Someone came.'

(2) **Kedagi** pole kodus.
someone.PRT *be.NEG* *home.INE*
'Nobody is home.'

(3) **Keegi** ei tea sellest.
someone *not* *know.CNG* *it.ELA*
'No one knows about this.'

(4) Ma ei tunne siin **kedagi.**
I *not* *know.CNG* *here* *someone.PRT*
'I don't know anybody here.'

In Standard Estonian, the indefinite pronouns *keegi* and *miski* 'something, anything, nothing' are differentiated by what they can refer to: *keegi* is strictly used to refer to animate entities, while *miski* refers to inanimate entities (Erelt 2017a:743). However, in some Estonian dialects, this distinction in animacy is not as clear, because *keegi* can also refer to inanimate entities, as in (5). In this paper, we aim to find out just how common such reference to inanimates is and how it is distributed geographically and functionally.

(5) Western (Martna)[2]
ei olnd **kedagi** poest saada
not *be.PST.CNG* *someone.PRT* *store.ELA* *get.INF*
'There was nothing to get from the store.'

A similar irregularity exists for the pronoun *kes* 'who' (see Pook 2019), but as with *kes*, the phenomenon is rarely mentioned in previous studies. In fact, only Viikberg (2020:174) mentions the possibility of *keegi* being used to refer to inanimate entities in the Mulgi dialect. Based on our previous research, however, this phenomenon exists in a much wider area than just that one dialect.

In this paper we regard animacy as a binary variable, following Fowler (1977: 16–17) in dividing and classifying as animate beings all those that are capable of initiating action and change and of movement. This means that all humans and animals are categorised as animate and everything else as inanimate. However, it must be acknowledged that typically animacy in language cannot be regarded as a binary variable at all, but rather as a scale from most to least animate. This scale is called the animacy hierarchy, which is presented by Dixon (1979:85) as follows:

> 1st, 2nd personal pronoun > 3rd personal pronoun > proper name > human noun > non-human animate noun > inanimate noun

For some languages or for some constructions, the distinction between these categories might be more fine-grained (e.g. having 1st and 2nd person as separate categories) or less fine-grained (e.g. only opposing animate to inanimate), but overall it is a universal tendency to grammatically distinguish those categories which are higher

in the hierarchy from those which are lower. Higher categories are often treated as more central to the clause structure and are more likely to act as an agent in events (Comrie 1989:185; Croft 1990:113; Whaley 1996:172; Kittilä, Västi & Ylikoski 2011:6).

The choice of treating animacy as binary in this paper stems from the nature of the data, which contain spoken texts on topics such as the informant's personal life, lifestyle, past events, or working methods, and where the marking of pronouns as biologically animate or inanimate was straightforward, i.e. without any borderline cases of animacy. Moreover, since this article studies the animacy of an indefinite pronoun, many of the finer categories in the animacy hierarchy cannot be applied to it at all.

This study has two aims. The first aim is to examine the data acquired from the Corpus of Estonian Dialects[3] and determine how *keegi* is used and what functions it fulfils in the dialects. This paper is a needed contribution to the field, as *keegi* (and most other indefinite pronouns in Estonian) and its use have never been thoroughly described before. As a continuation of previous research (see Pook 2019), the main aim of this paper is to study the use of *keegi* in regard to the animacy of its referent in order to ascertain which dialectal areas allow the variation of referring to both animate and inanimate entities with *keegi* and which variables influence this variation. The linguistic variables we use in our study help to explain under which conditions the inanimate *keegi* can be used. Our purpose is therefore to analyse this variation in spoken language and its relation to other relevant variables.

In addition, we aim to find out whether the geographical and morphosyntactic variables that affect the animacy-related use of the interrogative pronoun *kes* 'who', as shown in Pook (2019), are similar for the indefinite pronoun *keegi*. In a sense, we want to discern whether the reason why *keegi* may select only animate entities or both animate and inanimate entities is due to its interrogative component *kes*, which serves as a source of grammaticalisation for indefinite *keegi*. We expect that the non-selectivity between animate and inanimate referents is spread in the same dialect area for both *keegi* and *kes*, and that the choice between the use of animates and inanimates is conditioned at least partially by the same factors. As a working hypothesis we expect that the animacy distinction has less importance in the scope of negation, and consequently the use of *keegi* referring to inanimates occurs mostly when *keegi* functions as a negative polarity item.

This paper is structured as follows. In Section 2 we provide a brief overview of Estonian dialects and describe our dataset. In Section 3 we describe the use of Estonian indefinite pronouns and discuss the functions of the pronoun *keegi*. Section 4.1 explains our annotation system and Section 4.2 describes the statistical methods used in the analysis. Section 5 presents the results of the statistical analysis, while a discussion and our conclusions are included in Section 6.

2. Data

Estonian dialects are traditionally divided into 8–10 dialects and 105–120 subdialects. According to the latest classifications, the North Estonian dialect group includes the Insular, Western, Mid, and Eastern dialects, the Northeastern–Coastal dialect group is composed of the Coastal and Northeastern dialects, and the South Estonian dialect group consists of the Tartu, Mulgi, Võru, and Seto

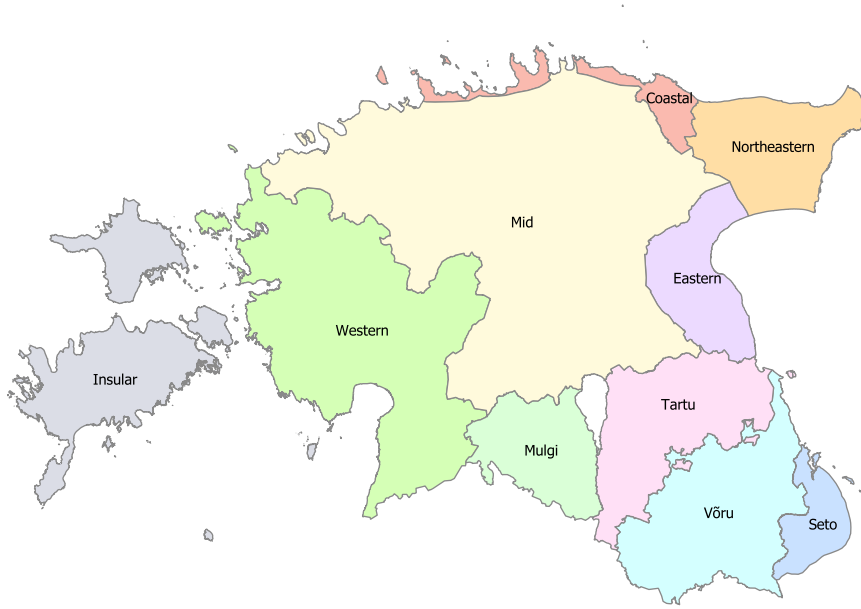

Figure 1. Estonian dialects.

dialects (Pajusalu 2007:231). This is the division used in the Corpus of Estonian Dialects and therefore also in this study (see Figure 1). It should be mentioned, however, that in earlier classifications the Northeastern and Coastal dialects were regarded as one dialect and the Seto dialect was considered to be a subdialect of Võru (Kask 1984). Every dialect is, in addition, divided into subdialects, which are based on the borders of historical parishes.

All the dialects are distinct from contemporary Standard Estonian, which is based on North Estonian but is also a compromise between various dialects, conscientious language planning, and recent influences of contact languages. Northern dialects share the most with Standard Estonian, with up to 58% common features (which include phonetic and grammatical features and core vocabulary) between the Mid dialect and Standard Estonian, while the southern dialects differ the most from Standard Estonian, with the Võru dialect sharing only 18% of common features with Standard Estonian (Pajusalu 2007:233).

The most significant differences in phonology, morphology, and lexis can be found between the southern and northern dialects, since South Estonian diverged from Proto-Finnic before other Finnic languages (Sammallahti 1977; Viitso 1985; Kallio 2012). However, recent dialect studies have found that on a (morpho) syntactic level, the biggest differences are between the eastern and western dialects instead, with the Coastal and Mulgi dialects fitting in with either group depending on the phenomenon studied (Lindström et al. 2009; Uiboaed 2013; Uiboaed et al. 2013; Lindström, Uiboaed & Vihman 2014; Lindström et al. 2015; Ruutma et al. 2016; Lindström & Uiboaed 2017; Lindström, Pilvik & Plado 2018; Pook 2021).

The data used in this study come from the Corpus of Estonian Dialects. The corpus contains authentic dialectal recordings from all dialect areas. The recordings

Table 1. The number of informants, total tokens, and lemma *keegi* in the data by dialect

Dialect	Informants	Total tokens	Lemma *keegi*
Western	54	251,031	476
Mid	72	246,167	595
Insular	37	202,325	138
Võru	22	111,503	122
Coastal	21	97,152	107
Eastern	19	48,353	128
Seto	17	68,414	63
Tartu	17	80,343	52
Mulgi	15	63,759	110
Northeastern	14	60,037	66
Total	**288**	**1,229,084**	**1,857**

are transcribed phonetically and annotated for morphological features. The speakers are typically older people, who have often lived in the same place their entire life and are therefore a good representation of their home dialect. The conversations cover a range of topics, such as their current lifestyle and family, past events, traditions, and working practices (Lindström, Lippus & Tuisk 2019).

This study uses the morphologically annotated texts, from which 1,857 observations of the pronoun *keegi* were compiled into our dataset. This also includes a few observations of the pronoun *kes* 'who' from the southern dialects, where *kes* (and its variants) have an indefinite meaning even without the affix *-gi*, as in (6). It has been claimed that previously the interrogative pronouns in Finno-Ugric languages were used for expressing indefiniteness; the *gi*-affixed forms are a later development in Finnic languages (Alvre 1986:49). Nowadays, the option to use interrogative pronouns indefinitely has receded from the written language, but can still be found in Votic, Veps, and in some Estonian and Finnish dialects (Alvre 1977:21, 1986: 46–49; Van Alsenoy & van der Auwera 2015:28; Karjalainen 2019).

(6) Võru (Rõuge)

tuulaiga	tiijäkki	iss	**kiä**	nakrist
that_time.ADE	*know.PST.CNG.CLI*	*not.PST*	*who*	*turnip.ELA*

'During that time nobody even knew about turnips.'

Table 1 gives an overview of the data used in this study.

3. The use of *keegi* and other indefinite pronouns

3.1 Indefinite pronouns

According to Martin Haspelmath's classic definition, indefinite pronouns are pronouns 'whose main function is to express indefinite reference' (Haspelmath

1997:11). However, as shown by Haspelmath himself and later by, for example, Denić, Steinert-Threlkeld & Szymanik (2022), indefinite pronouns may have various functions and various referential values, showing that indefiniteness is not a clear-cut category and is internally heterogeneous. Haspelmath (1997) has listed nine main functions of indefinite pronouns, and Denić et al. (2022) have reduced this number to six main semantic 'flavours': *specific known*, *specific unknown*, *nonspecific*, *negative polarity*, *free choice*, and *negative indefinite*. Most European languages have more than one indefinite pronoun for covering this range of meanings; however, in Estonian, *keegi* can be used for all of them.

Indefinite pronouns are very common within the scope of negation. Most European languages use special negative indefinite pronouns (Bernini & Ramat 1996:120), such as *nobody* in English. Estonian is one of the few European languages that does not have dedicated negative indefinites; only *mitte keegi* (which includes the non-sentential negation marker *mitte*) has grammaticalised into this function to a certain degree (Bernini & Ramat 1996:124–125). Negative indefinites may co-occur with verbal negation or themselves suffice to express sentential negation (as in English) (Haspelmath 1997:36). In Estonian, *mitte keegi* always occurs with verbal negation.

Another widely discussed function of indefinites in negative contexts is negative polarity. Negative polarity items are words or phrases that can be used only in sentences that include at least one negative element in the same sentence (Zwarts 1999:295). In relation to indefinite pronouns, well-known polarity items are the English *any*-series (*anybody*, *anything*). In addition to negative clauses they can be used in some other negative-polarity environments, such as in conditional or interrogative clauses, as well as some other environments, and are not strictly related to the expression of non-existence (Haspelmath 1997:37–39), thus in typical *irrealis* contexts. Estonian, again, does not have a dedicated indefinite pronoun for expressing negative polarity and also uses *keegi* in negative polarity contexts.

In many languages, however, indefiniteness can also be expressed in negative contexts by other means. Partee (2008) has explained the use of Russian partitive-genitive within the scope of negation by referring to decreased referentiality and non-veridicality in this context. Furthermore, based on Kiparsky (1998), Partee shows that the partitive marking of an object in Finnish occurs in a context of lowered referentiality (compared to the total object in the accusative). The connection between non-referentiality under the scope of negation and partitive marking of NPs with reduced referentiality has been found in many languages, but especially in Balto-Finnic and Slavic languages (Miestamo 2014; Seržant 2015). According to Seržant (2015), the partitive-under-negation rule is a language-contact phenomenon and common Eastern Circum-Baltic innovation. The use of partitive marking of objects and existential subjects under negation is obligatory in Estonian as well; it also applies to indefinite pronouns, e.g. *keegi* (nominative) > *kedagi* (partitive).

3.2 Indefinite pronouns in Estonian

While personal, demonstrative, and interrogative pronouns in Finno-Ugric languages are fairly old word classes, indefinite pronouns formed considerably later, as evidenced by their varied origins and the existence of compound forms (Alvre 1980:539, 1986:5).

Van Alsenoy and van der Auwera (2015:32, 39, 66) categorise Uralic indefinites into four groups: negative indefinites (morphologically negative), negative indefinites (morphologically non-negative), negative polarity indefinites, and neutral indefinites. Out of these four categories, Estonian mostly uses neutral indefinites, which do not have any distributional restrictions: even when used with a negative verb they acquire their negative or specific meaning from the context. This can result in ambiguity in meaning in some cases. However, Estonian also has a non-sentential negative marker *mitte* 'not', which, used together with *keegi* 'nobody' or *miski* 'nothing', has the function of emphasising the negativity and clarifying the meaning. In the previously mentioned categories, *mitte* + indefinite pronoun can be considered to be a morphologically negative indefinite, or a negative indefinite in terms of Haspelmath (1997) and Denić et al. (2022).

Interestingly, the word *mitte* is etymologically related to the partitive form of the interrogative *mis* 'what' (Mägiste 2000:1545). Since indefinites have developed from interrogatives in Estonian, the proposed development from **mitä-ä-hen* > *mittää* > *mitta* > *mitte* (Mägiste 2000:1545) indicates how tightly the use of interrogative-indefinite pronouns and partitive case marking are related to each other especially in negation contexts.

Moreover, *mitte* is also used as a constituent negator with infinitive and converb clauses (e.g. *mitte tea-des* not know-CONV 'not knowing') in Standard Estonian (see Tamm 2015), and as a negation word or polarity item in some dialects, especially in the Insular and Western dialects, as in (7). Thus, the use of interrogative/indefinite pronouns in the context of negation was also common in the past and it has developed into a polarity item and/or a negation word in Estonian.

(7) Western (Kullamaa)

pole	nüid	kellegile	kirjuttand	kaa	**mitte**
be.NEG	*now*	*someone.ALL*	*write.PST.PTCP*	*also*	*not*

'[I] haven't written to anyone at all now.'

It can be explained by the fact that the use of partitive case under the scope of negation is a common feature in Estonian as well as in other Finnic languages and in Baltic and Slavic languages; in these languages partitive marking is used for expressing indefinite, non-referential meanings (Miestamo 2014; Seržant 2015). Thus partitive indefinite pronouns are something that could be expected to occur in negated clauses (as a subject or object argument under the scope of negation), and therefore the development from a partitive indefinite pronoun to a polarity item and later into a negation word seems possible.

One of the most productive affixes for deriving indefinite pronouns is *-gi/ki*, which works in Estonian in a way similar to discourse particles and has various meanings related to information structuring, quantification, etc. (Metslang 2003). The original meaning of the affix *-gi/ki* is unclear; in present-day data it has both additive ('also') and scalar ('even') meanings. In negative contexts it behaves as a negative polarity item, as many words with this affix are used only with negative polarity (Sang 1983:121–122; Paldre 1998:49–51). It is possible that *-gi/ki* has become a part of many indefinite pronouns precisely through negative polarity.

The Estonian indefinite pronouns with the suffix *-gi/ki* are *keegi*, *miski*, *mingi* 'some, a certain', *kumbki* '(n)either', and *ükski* 'none'; the first four of these are based on early interrogative stems, the last one on the numeral *üks* 'one' (Alvre 1980:539; see also Nevis 1984). Deriving indefinites from interrogatives is common typologically (Haspelmath 2013) and is characteristic of the Uralic languages (Van Alsenoy & van der Auwera 2015). When looking at our dialectal data, only the South Estonian Võru and Seto varieties use bare interrogatives (without *-gi/ki*) as indefinites (*kiä* 'who, somebody').

Deriving indefinites with the *-gi/ki* clitic is thus a relatively late development, which can also be seen from the position of *-gi/ki*. As an enclitic particle, it is attached to the very end of the word after any number and case markers (*ilusa-te-le-gi* 'beautiful-PL-ALL-CLI'), but as an affix on indefinites its position varies: it is used before or after the case marker, e.g. *kelle-le-gi – kelle-gi-le* (see Pant 2018; Pant 2020). This positional variation is an indicator of the ongoing lexicalisation process, whereby the *-gi/ki* clitic becomes a part of the stem and therefore its natural position is before the case and number suffixes (*kellegi-le*). However, language planning still suggests the placement of *-gi/ki* after other suffixes, similarly to the use of the *-gi/ki* clitic as a discourse particle (Pant 2018). In dialects, the typical position of *-gi/ki* is before the case marker, at least in the allative form (Saareste 1955:16), and this does appear in our data: out of 35 allative forms, 23 have the case marker at the end, while 10 pronouns end with *-gi/ki* (and two pronouns from the Seto dialect lack a marker for indefiniteness).

Other indefinite pronouns in Estonian are *kõik* 'all', *iga* 'each', *mõlemad* 'both', *kogu* 'all', *mitu* 'many', *mõni* 'some', *üks* 'one', *teine* 'other', etc. (Erelt, Erelt & Ross 2007:187). The use of the pronouns *mingi* and *üks* has been more thoroughly examined by Pajusalu (2000, 2001, 2004): while both of these pronouns express vagueness in spoken language, using *mingi* leaves an impression that the referred entity is unfamiliar to both the speaker and the listener, while *üks* conveys the meaning that in that given context the referent is unknown only for the listener; *mingi* can also have a negative or evaluative connotation, while *üks* typically does not (Pajusalu 2000). It has been argued that indefinite pronouns such as *kõik*, *mõni*, and *mitu* should more accurately be called quantifying pronouns, as they are often used as definite pronominal NPs in spoken language (Pajusalu 2009:135).

3.3 *Functions of* keegi *in the data*

In this section we describe the possible functions that the pronoun *keegi* can have based on the data from the CED. The functions are defined on the basis of syntax. The indefinite pronoun can be used as an argument (subject, object, oblique argument), an attribute, a negative polarity item, and as some other minor functions that are mostly related to spoken use of language and are therefore not mentioned in Estonian grammars. We have broadly referred to all of these uses as functions of *keegi*. This categorisation is our own and does not follow any previously described functions for the pronoun *keegi*.

Nominative subject

The subject argument in Estonian is typically in the nominative case and agrees with the verb in person and in number (Erelt, Metslang & Plado 2017:240).

The indefinite pronoun *keegi* often occurs in subject position and indicates that the subject's referent is unknown or even irrelevant for the speaker and/or listener, as in (8).

(8) Western (Kirbla)
keegi oli ennustand vanal aeal
someone *be.PST.3SG* *predict.PST.PTCP* *old.ADE* *time.ADE*
'Someone had predicted in the old times.'

Partitive subject

Estonian has the option of using partitive subjects which alternate with nominative subjects, a case of differential subject marking (see e.g. de Hoop & de Swart 2009). The use of a partitive subject is more restricted than that of a nominative subject: a partitive subject occurs most commonly in existential and possessive clauses with XVS[4] word order, and is obligatory in negative existential (as in (9)) and possessive clauses (Erelt & Metslang 2006:255); in all of these clause types, it alternates systematically with a nominative subject. However, the use of a partitive subject is not limited only to these clause types (Huumo & Lindström 2014; Lindström 2017); its use is mostly linked to quantitative indefiniteness (Metslang 2012; Lindström 2017). Partitive subjects here are categorised separately from nominative subjects since *keegi* as a partitive subject behaves significantly differently from *keegi* as a nominative subject, as shown in the statistical analysis in Section 5.

(9) Insular (Kihelkonna)
nüid äp pole **kidad** kiss ära koristab
now *not* *be.CNG* *someone.PRT* *who* *away* *clean.3SG*
'Now there is nobody to clean [it] up.'

Object

Estonian has differential object marking, meaning that the marking of the direct object varies and is dependent on several semantic and syntactic factors (see e.g. Ogren 2015). The object is most typically marked with the partitive case (for partial objects) and with the genitive or nominative case (for total objects). The choice between using a partial or a total object is dependent on polarity, aspect, and the referent's boundedness. If a clause is perfective, the referent is quantitatively bounded, and the clause is affirmative, a total object is used. If even one of these conditions is not met, a partial object is used instead (Metslang 2017:258, 264–267). Some verbs, however, take only partitive objects and do not allow object marking alternations (see Tamm & Vaiss 2019). Interestingly, in the dataset of this study, all the objects are in the partitive case; 87% of them occur in a negative sentence.

(10) Eastern (Palamuse)
maa i oska **kedagi** enam kõnelda
I *not* *know.CNG* *someone.PRT* *more* *say.INF*
'I don't know what else to say.'

Adverbial

In the Estonian grammar tradition, the term *adverbial* covers both oblique arguments (such as arguments marking experiencer, possessor, or addressee) and adjuncts (e.g. time and location adverbials). The border between the oblique arguments and adverbials is not always clear-cut in Estonian: on one hand, the option to have an oblique argument and the form of it are selected by the predicate; on the other hand, their presence in the clause is far from being obligatory and is more likely context-dependent (see e.g. Lindström & Vihman 2017), making obliques closer to adjuncts. Therefore we use a cover term *adverbial* in this study, without drawing out clear differences between the obliques and adjuncts. In (11) *keegi* is an adjunct (semantically beneficiary), in (12) it is a possessor argument, and in (13) it is an addressee. Most of the uses of *keegi* in this group are related to the marking of possessors, addressees, and beneficiaries. Note that some typical adjuncts, such as locatives and time adverbials, cannot be formed with the indefinite pronoun *keegi*.

(11) Western (Kullamaa)
see joosis jälle oma tuast välla tegi värava
that_one run.PST.3SG again their room.ELA out do.PST.3SG gate.GEN
kellegil lahti
someone.ADE/ALL open
'That one ran out of their room again, opened the gate for someone.'

(12) Mulgi (Paistu)
nüid eij ole **kellekkil** külunõud äm
now not be.CNG someone.ADE sowing_vessel.PRT anymore
'Now no one has a sowing vessel anymore.'

(13) Mid (Peetri)
põle seda irmust muret et mul vel **kellegille**
be.NEG that.PRT terrible.PRT worry.PRT that me.ADE still someone.ALL
maksu maksta onn
tax.PRT pay.INF be.3SG
'I don't have that terrible worry that I still have to pay taxes to someone.'

Genitive attribute

A genitive attribute occurs within the NP and precedes the head noun. Estonian genitive attributes may express the possessor, author, place, time, quantum, purpose, etc. (Pajusalu 2017a:388). In our data, all the uses were more or less closely related to possessor marking, as in (14). Only the uses where the indefinite pronoun has the meaning 'proper, true' could be seen as a separate group, as in (15).

(14) Mid (Pilistvere)
ei tiantki keegi millas **kellegi** sünnibääv oli
not know.PST.CNG.CLI nobody when someone.GEN birthday be.PST.3SG
'Nobody knew when anyone's birthday was.'

(15) Eastern (Palamuse)
sa kurat põle **kellegi** miis
you devil be.NEG someone.GEN man
'You're no one's damn man.'

Postnominal attribute

Estonian has mostly prenominal attributes in noun phrases (e.g. genitive attributes), as they are strongly preferred over postnominal attributes, but postnominal attributes are also possible (Pajusalu 2017a:382). *Keegi* as a postnominal attribute typically belongs to a pronoun (*me* 'we', *nad* 'they', as in (16)) or to a noun referring to a group of people (e.g. *rahvas*, *inimesed* 'people'). This construction has the meaning 'any of the group' or 'none of the group'.

(16) Northeastern (Lüganuse)
me **keegi** ei kuuld seda
we *someone* *not* *hear.PST.CNG* *that.PRT*
'None of us heard that.'

Determiner

Since Estonian lacks grammatical articles, indefinite article-like determiners *keegi*, *miski* 'something, nothing', *üks* 'one', *mingi* 'some, a certain', etc. can be used to express indefiniteness. These determiners are more frequent in spoken than in written language (Pajusalu 2017a:382–384, 2017b:573). In this context, grammatically *keegi* can be replaced by *mingi* or *üks*, changing only minute nuances in the meaning (see Section 3.1), and *keegi* can be considered (as with *üks* and *mingi*) to function like an indefinite article (Pajusalu 2000:89), with the main function of indicating that the referent of the NP is unknown, as in (17).

(17) Mid (Suure-Jaani)
sial tulli **kiigi** sugulane
there *come.PST.3SG* *someone* *relative*
'There came a relative.'

Negative polarity item

A negative polarity item (NPI) is a word associated with a negation environment, which means it normally appears in sentences with negative polarity, but it is also common in certain non-negative contexts such as conditional or interrogative sentences. Typical NPIs in English are *any* (and the *any*-series), *ever*, *at all*, etc., although in different languages NPIs can range from nouns and adverbs to even verbs and constructions (Sang 1983:120; Haspelmath 1997:33–34; Giannakidou 2011:1661–1662; Erelt 2017b:193).

The affix *-gi* has been considered to be an NPI itself, as words like *ükski* 'none', *iialgi* 'never', *sugugi* '(not) at all', etc. are all used only with negative polarity. Although pronouns like *keegi*, *miski* 'something, nothing', *mingi* 'some, any' and adverbs like *kunagi* 'ever, never' have both positive and negative meanings, the first interpretation of their meaning in a negated sentence is negative exactly because of the affix *-gi* (Sang 1983:121–122; see also Paldre 1998). A study about negation in Estonian dialects found that *keegi* is used as an NPI in all of the analysed subdialects (the study included one subdialect from each dialect), but it was a more frequent means of emphasising negation in the subdialects of the Western, Mid, Eastern, and Mulgi dialects (Klaus 2009:148).

Since *keegi* can be used as a subject or object under negation and, based on its form, we cannot distinguish its use as a negative polarity item from other uses, we have taken a narrower approach to the definition of an NPI here: specifically, NPIs are those uses of indefinite pronouns in negated clauses that do not fill any argument position of the negated verb, i.e. their use is not related to the meaning of the main verb but only to the negation. NPIs in our data only appear in negative environments and have the purpose of emphasising the negation.

More than half of the NPIs in our data are preceded by *ega* 'nor', *ei* 'no', or *muud* 'other:PRT', as seen in (18), forming a somewhat grammaticalised construction. For the other NPIs, *keegi* typically acquires the meaning of 'at all', as seen in (19).

(18) Western (Martna)
täál ei olnd änam mingit peret ega **kedad**
s/he.ADE not be.PST:CNG anymore any.PRT family.PRT nor someone.PRT
'S/he didn't have a family or anyone anymore.'

(19) Mulgi (Helme)
raud ei kulu **kedägi**
iron not wear.CNG someone.PRT
'Iron doesn't wear out at all.'

Generalising alternative

In the data of this study, a generalising alternative follows an NP and refers to an indefinite, unspecified option similar to that NP (20). The NP in this structure is separated from the generalising alternative by *või/ehk* 'or', with the NP being in focus, while the following *või/ehk keegi* denotes uncertainty or possible other alternatives (Lindström 2001:96).

(20) Tartu (Otepää)
esä vai **keski** kes tulli siis jälle appi
father or someone who come.PST.3SG then again help.ILL
'Father or someone who came to help again.'

The distribution of the aforementioned functions in the data is depicted in Table 2. *Keegi* is most commonly used as a nominative subject and an object, followed by the functions of partitive subject, adverbial, and negative polarity item. *Keegi* is less often used as any type of attribute or as a generalising alternative.

4. Methods

4.1 Annotation

Our dataset consists of observations of *keegi* and its variants from the corpus. Each datapoint includes the preceding and following context (up to 20 words), the case marking of *keegi*, and information about the speaker. Each of the sentences in the dataset was manually annotated with the following variables.

Animacy of the referent

This is the dependent variable of the study and marks whether the entity that *keegi* is referring to is animate or inanimate. In this study, all humans (including human

Table 2. The frequency of the functions of *keegi* in the data

Function	Frequency
nominative subject	542 (29.2%)
object	435 (23.4%)
partitive subject	304 (16.4%)
adverbial	207 (11.1%)
negative polarity item	184 (9.9%)
determiner	102 (5.5%)
generalising alternative	32 (1.7%)
genitive attribute	26 (1.4%)
postnominal attribute	25 (1.4%)
Total	**1,857 (100%)**

collectives) as well as animals are marked as animate, and everything else is marked as inanimate. As mentioned previously, in real language use, animacy is a much more complex concept and not just a binary division, but in the interest of operationalisation, while also taking into account the topics and themes in the spoken data used, it is reasonable to differentiate only between animate entities, as in (21), and inanimate entities, as in (22).

(21) Võru (Rõuge)
niisama panõ õik **kiäkki** rätti pähäq
*without_reason put.*CNG *not someone headscarf.*PRT *head.*ILL
'No one puts a headscarf on without a reason.'

(22) Insular (Kihnu)
põlõ vanainimesel enäm tehä **kedäd**
*be.*NEG *old_person.*ADE *anymore do.*INF *someone.*PRT
'An old person has nothing to do anymore.'

Polarity of the clause

This marks whether the polarity of the clause containing *keegi* is affirmative, as in (23), or negative, as in (24). We predict that the animacy distinction has less importance within the scope of negation; therefore referring to inanimate entities with *keegi* could be more common in negative clauses.

(23) Coastal (Jõelähtme)
kiegi vanem mies läks Tallinna
*someone old.*CMP *man go.*PST.3SG *Tallinn.*ILL
'An older man went to Tallinn.'

(24) Mid (Kose)
mina i tia neist vanust juttudest **kedagi**
*I not know.*CNG *this.*PL.ELA *old.*PL.ELA *story.*PL.ELA *someone.*PRT
'I don't know anything about those old stories.'

Function of keegi

This marks which syntactic function *keegi* fills in a clause. These functions are as follows: nominative subject, partitive subject, object, adverbial, genitive attribute, postnominal attribute, determiner, negative polarity item, and generalising alternative. See Section 3.2 for a more detailed description of the functions.

Position of keegi *in the clause*

This marks one of three places in the clause for *keegi* to be situated: clause-initially, as in (25), clause-internally, as in (26), or clause-finally, as in (27).

(25) Insular (Jämaja)
kiskid oli ikka kodu
someone *be.PST.3SG* *always* *home.INE*
'Someone was always home.'

(26) Tartu (Nõo)
külän **kellegil** ess ole uibuaida
village.INE *someone.ADE* *not.PST* *be.CNG* *apple_orchard.PRT*
'No one in the village had an apple orchard.'

(27) Võru (Põlva)
mes sa keelä vai käsek **kedä**
what *you* *forbid.2SG* *or* *order.2SG* *someone.PRT*
'Why you forbid or order anyone.'

Case marking of keegi

This variable was extracted directly from the extant corpus annotation and marks the case of *keegi* in the clause. Out of the 14 Estonian cases, eight are found in the data: nominative, genitive, partitive, elative, allative, adessive, ablative, and comitative. In a previous animacy study of *kes* 'who', it was found that case was significantly associated with the referent's animacy, with elative and comitative being the most frequently used cases to refer to inanimate referents (Pook 2019), so it is highly likely that the case of *keegi* also affects its use.

Dialect

This marks which dialect area the speaker is from: the Coastal, Northeastern, Insular, Western, Mid, Eastern, Mulgi, Tartu, Võru, or Seto dialect. We predict that dialects are a very significant factor determining the probability of referring to an inanimate entity with *keegi*. In a previous study of *kes* 'who', the pronoun was used to refer to inanimate referents most frequently in the northern dialects, particularly in the Eastern, Western and Coastal dialects, while using *kes* in that manner was rare or unattested in the southern dialects (Pook 2019). We expect the area where *keegi* is used for inanimates to be roughly the same.

Table 3 gives an overview of all the variables used in this study.

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

Variable	Values
animacy of the referent (animacy)	animate
	inanimate
polarity of the clause (polarity)	affirmative
	negative
function of *keegi* (function)	nominative subject (nom_subject)
	partitive subject (part_subject)
	object
	adverbial
	genitive attribute (gen_attribute)
	postnominal attribute (post_attribute)
	determiner
	negative polarity item (pol_item)
	generalising alternative (gen_alternative)
position of *keegi* in the clause (position)	clause-initial (beginning)
	clause-internal (middle)
	clause-final (end)
case marking of *keegi* (case)	nominative (nom)
	genitive (gen)
	partitive (prt)
	elative (el)
	allative (all)
	adessive (ad)
	ablative (abl)
	comitative (com)
dialect	Eastern (EST)
	Western (WST)
	Mid (MID)
	Insular (INS)
	Coastal (CST)
	Northeastern (NE)
	Mulgi (MUL)
	Tartu (TRT)
	Võru (VRU)
	Seto (STO)

4.2 Statistical analysis

When studying dialect syntax, it is highly beneficial to have a large number of natural language recordings since it can be difficult to reproduce syntactic phenomena in a controlled environment. However, this type of data collection can also result in an unpredictably unbalanced dataset, in which the phenomenon of interest can be represented many times in one dialectal area or construction and hardly ever in another due to arbitrary and uncontrollable factors during data collection, but not necessarily due to the actual distribution of the phenomenon.

Hence, in this study, we have used three different statistical methods, none of which pose any particular requirements upon the data, making them highly suitable to use in the case of unbalanced datasets with categorical variables. Specifically, these methods are conditional inference trees, random forests, and multiple correspondence analysis. We applied all of these in order to determine which variables affect the use of *keegi* in referring to animate or inanimate entities.

Conditional inference trees and random forests are methods based on binary recursive partitioning. At each stage, the tree model's algorithm tests the association between the independent variables and the given response variable (which, in this study, is the animacy of the pronoun *keegi*). The variable most strongly associated with the response variable is the one used to split the data into two sets. This kind of partitioning continues until no variable is associated with the response at a level of statistical significance. At this point, the results are depicted as a tree with binary splits (Hothorn, Hornik & Zeileis 2006; Strobl, Malley & Tutz 2009).

For random forests, the model outputs a measure of importance for each variable, averaged over many conditional inference trees. These measures, in turn, reflect the value of impact each variable has on the response. The goal of these two methods is to predict the chances of the dependent variable occurring in a given context, specified by the independent variables (Breiman 2001).

Correspondence analysis (CA) is an exploratory technique designed specifically for the analysis of categorical variables. CA takes the frequency of co-occurring features and converts them to distances, which are then plotted on a two- or three-dimensional graph to visualise how the variable values are associated with each other (Glynn 2014:445). Multiple correspondence analysis is an extension of simple CA, but the former has the ability of analysing more than two factors simultaneously (Hill & Lewicki 2006:136).

All three of these methods have been successfully used in many other studies of Estonian, Estonian dialects and (dialect) syntax (see e.g. Uiboaed 2013; Ruutma et al. 2016; Lindström & Uiboaed 2017; Taremaa 2017; Lindström, Pilvik & Plado 2018; Pook 2019; Hint et al. 2021; Lindström, Pilvik & Plado 2021; Pook 2021).

All of 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 *party* package (Hothorn, Hornik & Zeileis 2006). The correspondence analysis was computed using the function *mjca()* from the *ca* package (Nenadic & Greenacre 2007).

Table 4. The frequency of animate and inanimate referents by dialect

Dialect	Animate	Inanimate	Total
Western	176	300 (63%)	476
Mid	251	344 (57.8%)	595
Eastern	56	72 (56.3%)	128
Insular	77	61 (44.2%)	138
Mulgi	67	43 (39.1%)	110
Northeastern	46	20 (30.3%)	66
Coastal	86	21 (19.6%)	107
Tartu	43	9 (17.3%)	52
Võru	122	– (0%)	122
Seto	63	– (0%)	63
Total	**987**	**870 (46.9%)**	**1,857**

5. Results

In this section we present our analysis of all the variables included in the study in terms of how they relate to *keegi* referring to animate and inanimate entities. In Section 5.1 we look at all the variables individually: dialect, function, case marking, polarity, and position. In Section 5.2 we show the conditional inference tree and random forest models in order to determine how these variables together affect the speaker's choice in referring to animate or inanimate entities with *keegi*. In Section 5.3 we use a multiple correspondence analysis to visualise the associations between all the variables on a two-dimensional graph.

5.1 Impact of the studied variables

Out of the 1,857 observations of *keegi* in the dataset, 987 referred to animate and 870 to inanimate entities. While in Standard Estonian *keegi* can only refer to animate beings, in dialects this restriction clearly does not always exist and *keegi* is used almost equally to refer to both animate and inanimate entities.

In order to find out which variables affect the use of *keegi* in terms of referring to animate or inanimate referents, in this section we analyse all of them in comparison to the animacy of the referent. The variables examined are dialect (and subdialects), function, case, polarity, and position.

5.1.1 Dialects and subdialects

First we compared the frequency of referring to inanimate entities in the dialects and subdialects. As can be seen in Table 4 and Figure 2, the dialects for which it is most probable to refer to inanimate entities with *keegi* are the Western, Mid, and Eastern dialects, where over half of the pronouns refer to an inanimate being.

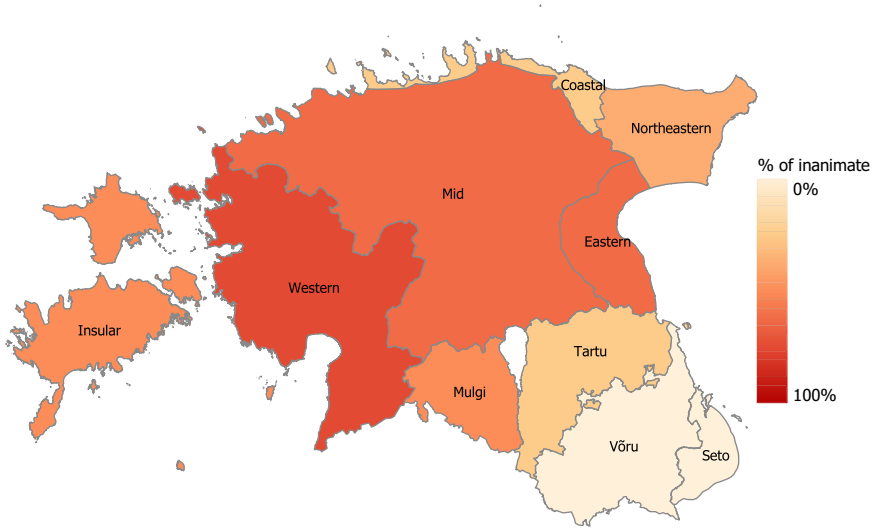

Figure 2. Percentage of the pronoun *keegi* used to refer to inanimate referents in dialects.

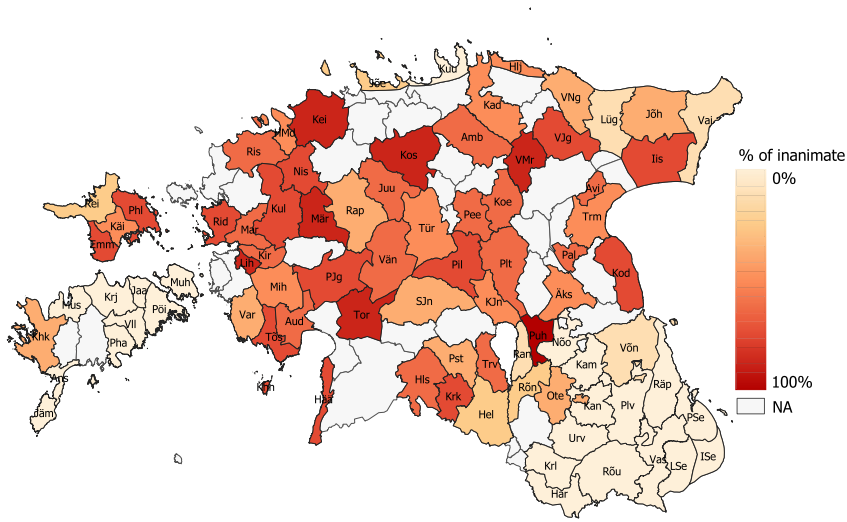

Figure 3. Percentage of the pronoun *keegi* used to refer to inanimate referents in the represented subdialects.

All in all, referring to inanimates is possible in all of the dialects except for the Võru and Seto dialects, where all of the instances refer to an animate being.

Looking more closely at the subdialects (Figure 3), we can see that most of the subdialects in the Western, Mid, and Eastern dialects have a high percentage of references to inanimate entities. The Insular dialect is split into two – although it has a moderately high probability of referring to inanimate beings with *keegi*,

Table 5. The percentage the pronouns *kes* and *keegi* used to refer to inanimate referents in dialects

Dialect	% of inanimate *kes*	% of inanimate *keegi*
Western	13%	63%
Mid	14.2%	57.8%
Eastern	14.4%	56.3%
Insular	6.5%	44.2%
Mulgi	10%	39.1%
Northeastern	5.7%	30.3%
Coastal	13.8%	19.6%
Tartu	5.8%	17.3%
Võru	2.4%	0%
Seto	0%	0%
Total	**9.7%**	**46.9%**

the data from the subdialects show that on the island of Saaremaa (the biggest island in the Insular dialect) most of the pronouns refer to animate entities, while on the island of Hiiumaa (the second largest island in the Insular dialect) it is very likely to refer to inanimate entities as well.

We compared the dialectal results obtained in this study about *keegi* with the results of the study about the use of the interrogative/relative pronoun *kes* 'who' (Pook 2019), which can also be used to refer to both animate and animate entities in Estonian dialects. Table 5 shows that the area where this variation occurs is quite similar. Although *kes* is predominantly used to refer to animate beings, with an average of only 9.7% of the pronouns referring to inanimates, the Western, Mid, and Eastern dialects have a higher percentage of inanimate referents, while the Võru and Seto dialects have few or no inanimate referents for both pronouns. The use of the pronouns in the Insular dialect is also divided in a similar manner between the islands of Saaremaa and Hiiumaa.

The significant differences in the percentages show, however, that while *kes* is mostly still perceived to be associated with animate entities, *keegi* has lost some of its distinction in animacy in the minds of the speakers and can be more easily used to refer to both animate and inanimate beings. Heine and Kuteva (2006:206, 227) have noted that, in many languages, as the interrogative markers have gone through the stages of grammaticalisation – from being just an interrogative marker to a marker that can introduce headed relative clauses – they have lost their distinction in gender, animacy, number, case, etc. Pook (2019) showed that this was also true for *kes*, as the pronoun was much more likely to refer to an inanimate entity when it was a relative pronoun than when it was used as an interrogative pronoun. Since indefinite pronouns have also grammaticalised from interrogatives, it is interesting to see that the semantic bleaching in animate–inanimate distinction is even more common with *keegi* than with *kes* (as can be inferred from frequency information).

Table 6. Percentage of the pronoun *keegi* used to refer to inanimate referents and the normalised frequencies (with base 10,000) of *keegi* and *miski* in the CED

Dialect	% of inanimates for *keegi*	Normalised frequency of *keegi*	Normalised frequency of *miski*
Western	63%	18.96	8.17
Mid	57.8%	24.17	6.5
Eastern	56.3%	26.47	5.38
Insular	44.2%	6.82	24.22
Northeastern	30.3%	10.99	26.32
Mulgi	39.1%	17.41	21.8
Coastal	19.6%	11.01	26.56
Tartu	17.3%	6.47	29
Võru	0%	10.94	32.55
Seto	0%	9.21	26.9

It is also interesting to note that (based on the data in the Corpus of Estonian Dialects) while *kes* 'who' and *mis* 'what' are both used to refer to both animate and inanimate entities in certain Estonian dialects, the same cannot be said about their counterparts *keegi* and *miski* 'something, nothing, anything', as *miski* can only refer to inanimate entities in Standard Estonian as well as in all the Estonian dialects.

We briefly examined the normalised frequencies of *keegi* and *miski* in the Corpus of Estonian Dialects to see whether the dialects that overwhelmingly use *keegi* to refer to both animate and inanimate entities therefore have a lower frequency of *miski* overall, since *keegi* fills the function of both pronouns, and whether the dialects that use *keegi* predominantly to refer to animate entities have a higher overall frequency of *miski*.

As can be seen in Table 6, our hypothesis is true for most dialects. The Western, Mid, and Eastern dialects, which have a very high percentage of *keegi* referring to inanimates, have a much lower usage frequency of *miski* than of *keegi*. Inversely, in the Coastal, Tartu, Võru, and Seto dialects, where references to inanimate entities using *keegi* are less common or even completely unattested, the frequency of *miski* is three or more times higher than the frequency of *keegi*.

The Insular dialect stands out because of its opposite behaviour: although almost half of the instances of *keegi* in that dialect refer to inanimates, the frequency of *miski* in the corpus is almost four times higher than the frequency of *keegi*. It is possible that there are other words used in a similar function and position, for example *üht(i)* '(not) at all' or *mitte* 'not' in the scope of negation. Previous researchers have also noticed the frequent occurrence of *miski* and *mitte* for emphasising negation in the Insular dialect (see e.g. Vitsberg 1958:27, 202).

Table 7. The frequency of animate and inanimate referents by function

Function	Animate	Inanimate	Total
polarity item	13	171 (92.9%)	184
object	59	376 (86.4%)	435
generalising alternative	9	23 (71.9%)	32
partitive subject	88	216 (71.1%)	285
determiner	33	71 (68.3%)	104
postnominal attribute	23	2 (8%)	22
adverbial	197	8 (3.9%)	205
nominative subject	539	3 (0.6%)	561
genitive attribute	26	– (0%)	26
Total	**987**	**870 (46.9%)**	**1,857**

5.1.2 Function

Next we looked at all the functions of *keegi* in comparison to the animacy of what *keegi* was referring to (see Table 7). The nominative subject (see (28)), attributes, and adverbials stand out as they are rarely or never used to refer to inanimate entities. *Keegi* as a polarity item, object, or partitive subject is, however, used predominantly to refer to inanimate referents. Generalising alternatives and determiners are also more likely to be inanimate.

(28) Western (Häädemeeste)
Interviewer:
aga miss selle vastu aittab siiss kaa
but what that.GEN against help.3SG then also
'What helps against that then?'
Speaker:
ei aitta **keegi**
not help.CNG someone
'Nothing helps.'

5.1.3 Case marking

For indefinite *keegi*, case seems to be strongly associated with the referent's animacy, as can be seen from Table 8. Partitive stands out as the typical case used to refer to inanimate referents, with 83.7% of partitive pronouns referring to inanimate beings. Meanwhile, nominative, adessive, allative, and genitive are strongly associated with referring to animate beings. These percentages correspond well to the results in the previous section since subjects and objects showed similar probabilities of animate/inanimate references relative to their prototypical cases, nominative and partitive.

The rest of the cases have too few observations in the dataset to draw any clear conclusions about their use in this variation.

For the pronoun *kes*, case was also a significant factor determining whether the pronoun was used to refer to animate or inanimate entities. However, for *kes* the

Table 8. The frequency of animate and inanimate referents by case

Case	Animate	Inanimate	Total
elative	–	6 (100%)	6
partitive	165	848 (83.7%)	1,013
comitative	4	2 (33.3%)	6
genitive	39	3 (7.14%)	42
nominative	587	10 (1.7%)	597
adessive	156	1 (0.6%)	157
ablative	1	– (0%)	1
allative	35	– (0%)	35
Total	**987**	**870 (46.9%)**	**1,857**

Table 9. The frequency of animate and inanimate referents by polarity

Polarity	Animate	Inanimate	Total
negative	539	775 (59%)	1314
affirmative	448	95 (17.5%)	543
Total	**987**	**870 (46.9%)**	**1,857**

elative and comitative cases were the ones where the majority of pronouns were used to refer to inanimates (see Pook 2019).

5.1.4 Polarity

The speaker's choice of using *keegi* to refer to inanimate entities is also affected by the polarity of the clause. Table 9 shows that in clauses with negative polarity, it is much more likely that *keegi* refers to inanimate beings (59%) than in affirmative clauses (17.5%). Thus the restriction that *keegi* has an animate referent does not hold up well at all in negative clauses.

5.1.5 Position

Finally, using *keegi* to refer to inanimates is particularly probable if *keegi* is situated at the end of the clause as opposed to the beginning (see Table 10). This is most likely associated with the function *keegi* serves in the clause, as functions that encourage referring to inanimate entities are either overwhelmingly (in the case of partitive subjects and objects) or always (in the case of polarity items) in the middle or at the end of the clause.

5.1.6 Summary of the variables' effects on the referent's animacy

Looking at all these variables separately, we can say that *keegi* is mostly used to refer to inanimate entities in the Western, Mid, and Eastern dialects, in negative clauses,

Table 10. The frequency of animate and inanimate referents by position in the clause

Position	Animate	Inanimate	Total
end	174	483 (73.5%)	657
middle	697	363 (34.3%)	1,060
beginning	116	24 (17.1%)	140
Total	**987**	**870 (46.9%)**	**1,857**

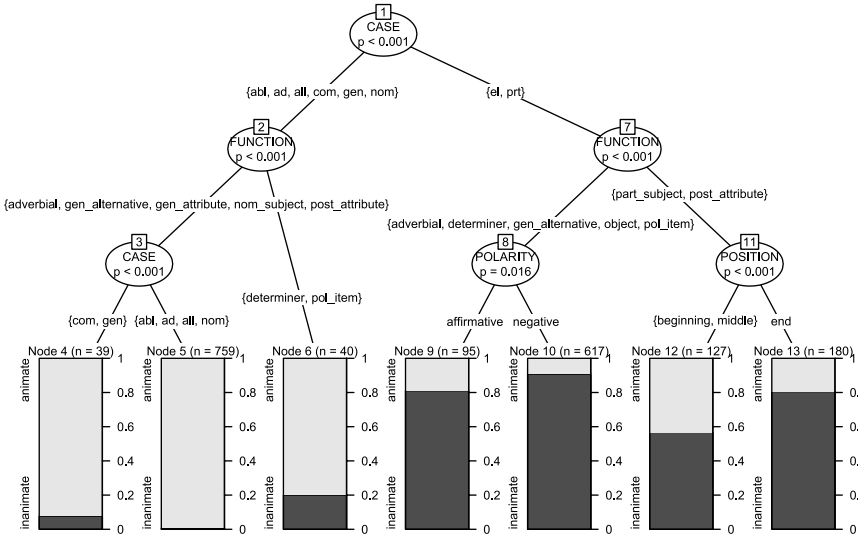

Figure 4. Conditional inference tree for the animacy of the entity that *keegi* is referring to.

as an object, a partitive subject, or a polarity item, and towards the end of the clause, as illustrated in (29).

(29) Mid (Koeru)
kõik pallajalu ühelgi põld jalas **kedagi**
everyone barefoot no_one be.NEG foot.INE someone.PRT
'Everyone [was] barefoot, no one had anything on their feet.'

In order to further verify these results, we have used multifactorial statistical methods in the next sections of this paper, which also give us the opportunity to measure the relations and interactions between the studied variables.

5.2 Conditional inference tree and random forest

In order to assess the significance of all the variables in association with each other, we ran a conditional inference tree model on the data. Figure 4 shows the

conditional inference tree graph for the animacy of the referent of the pronoun *keegi*. Here we focus on linguistic/functional variables only: the variables included in this model were case, function, polarity, and position; the response in this model was animacy. We have excluded the variable of dialect from this model, as its effect on the animacy of the referent has already been demonstrated in Section 5.1.1. Data from all dialects are still included in the analysis.

The figure displays all the possible splits significant at the level of 0.05 or less. The bar plots at the bottom show the proportion of animate (light grey) and inanimate (dark grey) observations with the given combination of variable values.

It can be seen that the animacy of *keegi* is significantly associated with all four included variables: case, function, position, and polarity. Case is the variable to first split the dataset into two: *keegi* in elative and partitive has a higher probability of being inanimate than *keegi* in other cases included in the data. Raw data show, however, that there are only six instances of *keegi* in elative in the dataset, which is definitely not enough to make any solid conclusions, so it should rather be said that partitive is the only case in the dataset that clearly licenses the inanimate use of the pronoun. The other cases are next divided by function: determiners and polarity items have a 20% chance of being inanimate (Node 6). The rest of the functions are split again by case: comitative and genitive have a low possibility of referring to an inanimate entity (Node 4), while ablative, adessive, allative, and nominative almost exclusively refer to animate beings (Node 5).

The set of partitive and elative is also split by function: adverbials, determiners, general alternatives, objects, and polarity items have a very high chance of referring to inanimate beings. If *keegi* in one of those functions is in a negative clause, the probability of it referring to an inanimate entity is even higher (Node 10) than when it is in an affirmative clause (Node 9). Partitive subjects and postnominal attributes behave according to their position in the clause: clause-initial or clause-internal *keegi* is less likely to refer to an inanimate entity (Node 12) than clause-final *keegi* (Node 13).

The C-index of concordance for this model is 0.94. The C-index evaluates the predictions made by the algorithm: it shows the proportion of concordant pairs divided by the total number of possible evaluation pairs. A value of 0.5 means that the model is not able to discriminate between the variants at all, a value between 0.5 and 0.7 shows poor discrimination, a value between 0.7 and 0.8 suggests an acceptable discrimination, a value between 0.8 and 0.9 shows excellent discrimination, and any value above 0.9 means that the model is able to discriminate between different variants exceptionally well (Hosmer Jr., Lemeshow & Sturdivant 2013: 177). Therefore this model is fitted exceedingly well.

While the conditional inference tree shows the significant associations between independent variables and the response, it does not show the strength of those associations. Therefore the random forest model was applied to the same dataset. This analysis includes the same variables as the conditional inference tree model, with the addition of the variable of dialect. The impact of the variables is shown in Figure 5. The names on the *y*-axis show the variables included in the analysis, and the numbers on the *x*-axis show the relative difference between the importance of the variables.

Figure 5 concludes that the most important predictor for the animacy of the pronoun *keegi* is dialect (0.042), followed by case (0.021), function (0.007) and

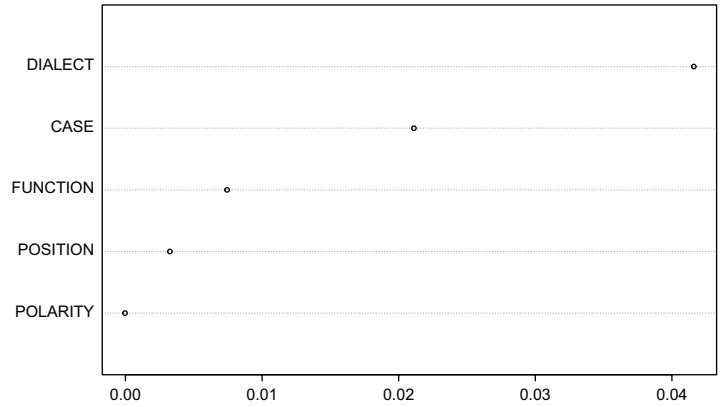

Figure 5. Random forest for the animacy of the entity that *keegi* is referring to.

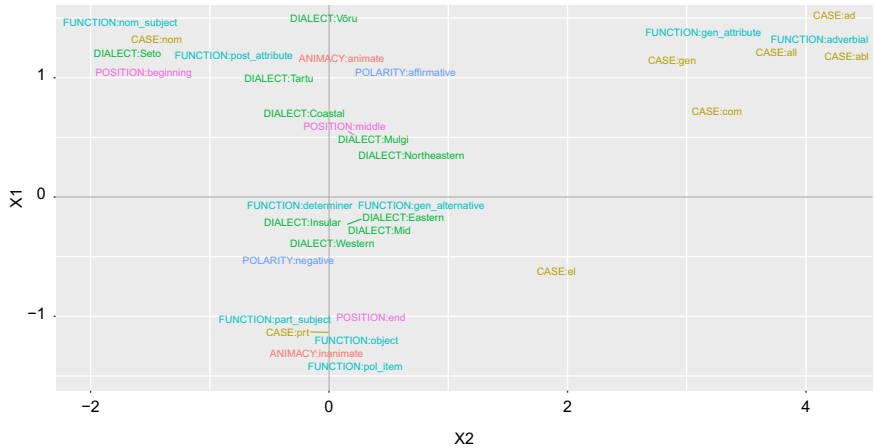

Figure 6. Multiple correspondence analysis for all the variables included in the data.

position of *keegi* (0.003). Polarity does not seem to have any discriminatory power in this model. The C-index of this model is 0.97, which suggests an outstanding fit.

This mostly reflects the results of the conditional inference tree, showing that the variables of case, function, and position affect the animacy of the pronoun both significantly and strongly. However, while polarity significantly determines the animacy of the pronoun in a certain context in the dataset, the association between polarity and animacy is weak and it cannot be generalised for the entire dataset.

5.3. Correspondence analysis

As a final method, we visualised all the studied variables with a multiple correspondence analysis (MCA) in Figure 6. For most datasets, the combination of the first

two dimensions offers the most accurate and easily interpretable visualisation of how the variables and their values are associated with each other (Glynn 2014:447). The further a value is from the origin (the point where the *x*-axis and *y*-axis intersect), the more discriminating it is. Inversely, the closer a value is to the origin, the less discriminating it is, but only in the context of the chosen variables. This means that a variable or a value might still contribute to the studied variation, but not in the visualised dimensions.

To analyse the relationship between one variable's value and another variable's value, one should look at the angle connecting the two values via the origin: the smaller the angle, the stronger the positive association probably is. If the angle is 90 degrees, the values are most likely not associated at all, and if the angle is 180 degrees, the values are probably negatively associated with each other.

It is important to note here that the MCA does not show whether the associations between the variable values are significant or relevant at all since the primary purpose of this technique is to just produce a simplified representation of the data. Therefore one must check all conclusions made with the MCA using raw data (Greenacre 1984:10; Hill & Lewicki 2006:134; Glynn 2014:444).

We can see in Figure 6 that the first, vertical dimension appears to be a continuum from most animate to least inanimate and describes 73.9% of the variance in the data. This dimension also mostly seems to follow the argument-marking schema, where prototypical nominative subjects are most likely to be animate (referring to a person), while objects and partitive subjects tend to refer to inanimates (concrete or abstract entities, events, or even non-referential use (see Metslang 2014:202)). It is not as obvious what the second, horizontal dimension represents. However, it only describes another 14.4% of the variance, so the vertical dimension is plainly much more important in describing the use of *keegi*. Combined, the first two dimensions describe 88.3% of the variance. This means that only 11.7% of the variance of these studied variables is left unexplained by this MCA analysis.

In addition to objects, inanimacy is also linked to partitive subjects and polarity items, to the partitive case in general and to clause-final position. These variable values are, however, not only associated with inanimacy, but many of them are also associated with each other. All objects and partitive subjects and a majority of polarity items are in the partitive case. Polarity items, in turn, typically occur at the end of the clause. Although negative polarity is situated a bit farther from the centre of this group, all three of the aforementioned functions are typically in the scope of negation and are associated with each other through that characteristic as well.

In fact, in some types of sentences, it can be somewhat difficult to make the distinction between objects, partitive subjects, and polarity items. See for example (30): in this clause, *häda* 'problem' could be interpreted as a partitive subject, making *kedagi* 'someone:PRT' a polarity item. However, if we consider *häda olema* 'to be wrong (with something)' to be a lexicalised verb construction, *kedagi* instead becomes the partitive subject of the clause.

(30) Mid (Ambla)

ei	old	tall	äda	**kedagi**
not	*be.PST.CNG*	*s/he.ADE*	*problem.PRT*	*someone.PRT*

'There was nothing wrong with him.'

Therefore, as it is sometimes difficult even to distinguish these three functions from each other, it is not at all unusual that they also function in a similar fashion in this variation, and the differences between them are more vague for the pronoun *keegi*. All in all, it is a cluster of values that truly function as a group, and none of them can be disregarded in analysing the use and variation of *keegi*.

Another group of associated values is the subject, clause-initial position, and the nominative case. These values are not as strongly linked to animates as the previously discussed values were to inanimates, as they are farther from each other on the plot and the angle connecting them to the origin is wider. Nevertheless, it is safe to say that this is another important cluster in describing the variation of *keegi*. While negative polarity is, in the given context of variables, not as discriminating in describing this variation, affirmative polarity is in fact very closely related to animate entities. Similarly, while none of the dialects are very strongly associated with inanimacy, the southern dialects and the Coastal dialect are clearly more connected to referring to animate entities.

A separate group is formed with the adverbial and genitive attribute functions and with genitive, ablative, allative, comitative, and adessive cases. Based on their position on the graph, it seems that both functions tend to be associated with animate entities. This is confirmed by the raw data, as there are only eight adverbials and no genitive attributes that refer to inanimate beings. As the name suggests, genitive attributes are all in the genitive case, while the rest of the mentioned cases are typically associated with adverbials, which explains why exactly these values are presented together on the graph.

All in all, this MCA analysis nicely illustrates the results obtained in the previous parts of the analysis: there are several significant variables in this study that all affect the use of *keegi*, and they do this in association with each other.

6. Conclusions and discussion

In this paper we examined the use of the indefinite pronoun *keegi* 'someone, nobody, anybody' in Estonian dialects. We described functions and positions in which *keegi* can be used in these dialects and analysed the phenomenon of using the otherwise animate *keegi* to refer to inanimate entities as well, a variation that is characteristic only of dialects and not of Standard Estonian.

Based on the data in the Corpus of Estonian Dialects, the pronoun *keegi* is used in the following functions: as a nominative and a partitive subject, an object, an adverbial, a genitive and a postnominal attribute, a determiner, a negative polarity item and a generalising alternative. Almost half of all the uses are subjects, but objects, adverbials, and negative polarity items are also very frequent.

The results show that *keegi* is most often used to refer to inanimate entities in the Western, Mid, and Eastern dialects, where over half of *keegi* pronouns refer to inanimates. At the same time, in the Võru and Seto dialects it does not seem to be at all possible to use *keegi* to refer to inanimate beings. Similar results were obtained in the study of the pronoun *kes* 'who' (Pook 2019), where it was possible to refer to inanimate entities with *kes* in the northern dialects, but this variation was rare or nonexistent in the Võru and Seto dialects. The Insular dialect's two biggest islands were also divided similarly in both studies – both *keegi* and *kes* can be used to refer to

inanimates on Hiiumaa, but rarely or never on Saaremaa. The similar distribution of inanimate uses of *kes* and *keegi* shows us that such developments are probably not coincidental: in this area the animate–inanimate distinction has for some reason started to fade.

Nevertheless, we cannot draw a direct line between the use of *kes* and the use of *keegi* in this similar variation. While the region where the speakers do not distinguish between animate and inanimate clearly overlaps for *kes* and *keegi*, the same cannot be said about their morpho-syntactic use. As our results show, the indefinite pronoun is most often used to refer to inanimate entities when *keegi* is an object, a partitive subject, or a negative polarity item, when it is in the partitive case, and positioned at the end of a negative clause. When *keegi* refers to an animate being, it is most likely a nominative subject at the beginning of an affirmative clause. In terms of *kes*, negation does not have a strong influence on this variation, and – contrary to *keegi* – the percentage of inanimate *kes* pronouns is three times higher in affirmative clauses than in negative clauses. In addition, instead of the partitive marking of *keegi* being the one most likely to refer to inanimates, it is the elative and comitative forms of *kes* that show the most prevalent lack of distinction in animacy.

However, it was shown in Pook (2019) that the distinction in animacy for the pronoun *kes* was most prevalent when *kes* was used as a relative pronoun, as opposed to an interrogative pronoun, that is, the more grammaticalised functions also showed the least selectivity in terms of animacy. A similar connection can be made for *keegi*, as the most grammaticalised function of a negative polarity item also increased its non-selectivity. So there are certainly parallels between the use of *kes* and *keegi*, but it is obviously not only due to the interrogative component *kes* in *keegi* that causes this variation.

Our results show how tightly indefinite pronouns and partitive case marking are interrelated in the scope of negation, as well as how the animate–inanimate distinction has become irrelevant in this specific context. From this we may also infer that we are dealing with a case of grammaticalisation: the loss of semantic distinctions or semantic bleaching more widely can be an early stage in the grammaticalisation process (see e.g. Heine & Kuteva 2002:2, 2006:60–61). The same has happened in the process of grammaticalisation of interrogatives into relative pronouns in many European languages (Heine & Kuteva 2006:209), including in Estonian (Pook 2019). The loss of a semantic distinction for indefinites, which is in the current case an extension of the inanimate uses of *keegi*, can be seen as an analogous grammaticalisation process, which can potentially result in developing into a negation word or a polarity item. We have seen that already happen with the word *mitte* 'not', which has grammaticalised from an interrogative/indefinite pronoun (in the partitive case) to a polarity item and/or negation word in Estonian (Mägiste 2000:1545). Thus the inanimate use of *keegi* in Estonian dialects seems to be following the same path of grammaticalisation, and does not seem to affect the animacy distinction very much in syntactic positions that are outside the scope of negation and for which differentiating between animate and inanimate referents is still relevant to understanding the content of the clause, such as for nominative (canonical) subjects or attributes.

Acknowledgements. This research has been supported by the Centre of Excellence in Estonian Studies (European Regional Development Fund). We want to thank the anonymous reviewers of the *Nordic Journal of Linguistics* for their valuable comments and suggestions.

Notes

1 Abbreviations follow the Leipzig Glossing Rules (2015): 1, 2, 3 = first, second, third person; ADE = adessive; ALL = allative; CLI = clitic; CMP = comparative; CNG = connegative; ELA = elative; GEN = genitive; ILL = illative; IMP = imperative; INE = inessive; INF = infinitive; IPS = impersonal voice; NEG = negative; PL = plural; PRT = partitive; PST = past tense; PTCP = participle; SG = singular.
2 This and all the following examples are derived from the Corpus of Estonian Dialects. Every example is preceded by the dialect, with the subdialect in parentheses.
3 https://doi.org/10.15155/1-00-0000-0000-0000-00076L (accessed 25 January 2019).
4 XVS stands for the dependent of the verb (X), the verb (V), and the subject (S).

References

Alvre, Paul. 1977. Pronoomenite morfoloogiat. Pronoomen *kes* [Morphology of pronouns: The pronoun 'who']. *Keel ja Kirjandus* [Language and literature] **1**, 18–26.
Alvre, Paul. 1980. *gi*-liitelisist pronoomeneist [On *gi*-joined pronouns]. *Keel ja Kirjandus* [Language and literature] **9**, 539–542.
Alvre, Paul. 1986. Läänemeresoome pronoomenite tüpoloogiat [Typology of Balto-Finnic pronouns]. *Fenno-Ugristica* **13**, 5–20.
Bernini, Giuliano & Paolo Ramat. 1996. *Negative Sentences in the Languages of Europe: A Typological Approach*. Berlin & New York: Mouton de Gruyter.
Breiman, Leo. 2001. Random forests. *Machine Learning* **45**(1), 5–32.
Comrie, Bernard. 1989. *Language Universals and Linguistic Typology: Syntax and Morphology*. Oxford: Blackwell.
Croft, William. 1990. *Typology and Universals*. Cambridge: Cambridge University Press.
Denić, Milica, Shane Steinert-Threlkeld & Jakub Szymanik. 2022. Indefinite pronouns optimize the simplicity/informativeness trade-off. *Cognitive Science* **46**(5), e13142.
Dixon, Robert M. W. 1979. Ergativity. *Language* **55**(1), 59–138.
Erelt, Mati. 2017a. Komplekslause. Liitlause [Complex sentences: Compound sentences]. In Mati Erelt & Helle Metslang (eds.), 647–755.
Erelt, Mati. 2017b. Lauseliikmed. Öeldis [Sentence constituents: Predicate]. In Mati Erelt & Helle Metslang (eds.), 93–239.
Erelt, Mati, Tiiu Erelt & Kristiina Ross. 2007. *Eesti keele käsiraamat* [Handbook of the Estonian language], 3rd updated edn. Tallinn: Eesti Keele Sihtasutus [Estonian Language Foundation].
Erelt, Mati & Helle Metslang. 2006. Estonian clause patterns: From Finno-Ugric to Standard Average European. *Linguistica Uralica* **42**(4), 254–266.
Erelt, Mati & Helle Metslang (eds.) 2017. *Eesti keele süntaks* [Syntax of the Estonian language] (Eesti Keele Varamu [Estonian language archive] III). Tartu: Tartu Ülikooli Kirjastus [University of Tartu Press].
Erelt, Mati, Helle Metslang & Helen Plado. 2017. Lauseliikmed. Alus [Sentence constituents: Subject]. In Mati Erelt & Helle Metslang (eds.), 240–257.
Fowler, Roger. 1977. *Linguistics and the Novel*. London: Routledge.
Giannakidou, Anastasia. 2011. Negative and positive polarity items. In Klaus von Heusinger, Claudia Maienborn & Paul Portner (eds.), *Semantics: An International Handbook of Natural Languge Meaning*, vol. 2, 1660–1712. De Gruyter Mouton.
Glynn, Dylan. 2014. Correspondence analysis: Exploring data and identifying patterns. In Dylan Glynn & Justyna Robinson (eds.), *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*, 343–485. Amsterdam: John Benjamins.
Greenacre, Michael J. 1984. *Theory and Applications of Correspondence Analysis*. London: Academic Press.
Haspelmath, Martin. 1997. *Indefinite Pronouns* (Oxford Studies in Typology and Linguistic Theory). Oxford: Oxford University Press.

Haspelmath, Martin. 2013. Indefinite pronouns. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/46.

Heine, Bernd & Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge University Press.

Heine, Bernd & Tania Kuteva. 2006. *The Changing Languages of Europe* (Oxford Linguistics). Oxford & New York: Oxford University Press.

Hill, Thomas & Pawel Lewicki. 2006. *Statistics: Methods and Applications: A Comprehensive Reference for Science, Industry, and Data Mining*. Tulsa: StatSoft.

Hint, Helen, Piia Taremaa, Maria Reile & Renate Pajusalu. 2021. Demonstratiivpronoomenid ja -adverbid määratlejatena. Miks me oleme siin ilmas, selles olukorras? [Demonstrative pronouns and adverbs as determiners: Why are we in the world, in this situation?] *Eesti ja soome-ugri keeleteaduse ajakiri* [Journal of Estonian and Finno-Ugric Linguistics] **12**(1), 79–111.

Hoop, Helen de & Peter de Swart (eds.). 2009. *Differential Subject Marking* (Studies in Natural Language and Linguistic Theory 72). Dordrecht: Springer Netherlands.

Hosmer Jr., David W., Stanley Lemeshow & Rodney X. Sturdivant. 2013. *Applied Logistic Regression*, 3rd edn. New York: John Wiley.

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

Huumo, Tuomas & Liina Lindström. 2014. Partitives across constructions: On the range of uses of the Finnish and Estonian 'partitive subjects'. In Silvia Luraghi & Tuomas Huumo (eds.), *Partitive Cases and Related Categories*, 153–176. De Gruyter Mouton.

Kallio, Petri. 2012. The non-initial-syllable vowel reductions from Proto-Uralic to Proto-Finnic. In Tiina Hyytiäinen, Lotta Jalava, Janne Saarikivi & Erika Sandman (eds.), *Per Urales ad Orientem: Iter polyphonicum multilingue. Festskrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februari 2012* [Through the Urals to the East: A polyphonic and multilingual journey. Festschrift dedicated to Juha Janhunen on his sixtieth birthday, 12 February 2012], 163–175. Helsinki: Suomalais-ugrilainen Seura [Finno-Ugrian Society].

Karjalainen, Heini. 2019. Borrowing morphology: The influence of Russian on the Veps system of indefinite pronouns. In Sofia Björklöf and Santra Jantunen (eds.), *Multilingual Finnic: Language Contact and Change* (Uralica Helsingiensia 14), 55–87. Helsinki: Suomalais-ugrilainen Seura [Finno-Ugrian Society].

Kask, Arnold. 1984. *Eesti murded ja kirjakeel* [Estonian dialects and written language]. Tallinn: Valgus.

Kiparsky, Paul. 1998. Partitive case and aspect. In Miriam Butt & Wilhelm Geudel (eds.), *The Projection of Arguments: Lexical and Compositional Factors*, 265–307. Stanford: CSLI Publications.

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

Klaus, Anneliis. 2009. *Eitus eesti murretes* [Negation in Estonian dialects]. Unpublished MA thesis, University of Tartu.

Leipzig Glossing Rules. 2015. Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses. https://www.eva.mpg.de/lingua/resources/glossing-rules.php (accessed 29 August 2022).

Lindström, Liina. 2001. Grammaticalization of *või/vä* questions in Estonian. In Ilona Tragel (ed.), *Papers in Estonian Cognitive Linguistics*, 90–118. University of Tartu.

Lindström, Liina. 2017. Partitive subjects in Estonian dialects. *Eesti ja soome-ugri keeleteaduse ajakiri* [Journal of Estonian and Finno-Ugric Linguistics] **8**(2), 191–231.

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

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

Lindström, Liina, Mervi Kalmus, Anneliis Klaus, Liisi Bakhoff & Karl Pajusalu. 2009. Ainsuse 1. isikule viitamine eesti murretes [Referring to the 1st person singular in Estonian dialects]. *Emakeele Seltsi aastaraamat* [Estonian Mother Tongue Society Year Book] 54, 159–185.

Lindström, Liina, Pärtel Lippus & Tuuli Tuisk. 2019. The online database of the University of Tartu Archives of Estonian Dialects and Kindred Languages and the Corpus of Estonian Dialects.

In S. Björklöf & S. Jantunen (eds.), *Multilingual Finnic: Language Contact and Change* (Uralica Helsingiensia 14), 327–350. Helsinki: Suomalais-ugrilainen Seura [Finno-Ugrian Society].

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

Lindström, Liina, Maarja-Liisa Pilvik & Helen Plado. 2021. Variation in negation in Seto. *Studies in Language: International Journal Sponsored by the Foundation 'Foundations of Language'* **45**(3), 557–597.

Lindström, Liina, Maarja-Liisa Pilvik, Mirjam Ruutma & Kristel Uiboaed. 2015. Mineviku liitaegade kasutusest eesti murretes keelekontaktide valguses [On the use of compound past tenses in Estonian dialects in the light of language contact]. *Võro Instituudi toimõtisõq* **29**, 39–70.

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

Mägiste, Julius. 2000. *Estnisches etymologisches Wörterbuch* [Estonian etymological dictionary], vol. 5, 2nd edn. Helsinki: Finnisch-Ugrische Gesellschaft.

Metslang, Helena. 2012. On the case-marking of existential subjects in Estonian. *SKY Journal of Linguistics* **25**, 151–204.

Metslang, Helena. 2014. Partitive noun phrases in the Estonian core argument system. In Silvia Luraghi & Tuomas Huumo (eds.), *Partitive Cases and Related Categories* (Empirical Approaches to Language Typology 54), 177–255. Berlin & Boston: De Gruyter Mouton.

Metslang, Helle. 2003. *-ki/gi*, *ka* ja nende soome kaimud [*-ki/gi*, *ka* and their Finnish cousins]. *Lähivertailuja* **12**, 57–81.

Metslang, Helle. 2017. Lauseliikmed. Sihitis [Sentence constituents: Object]. In Mati Erelt & Helle Metslang (eds.), 258–277.

Miestamo, Matti. 2014. Partitives and negation: A cross-linguistic survey. In Silvia Luraghi & Tuomas Huumo (eds.), *Partitive Cases and Related Categories* (Empirical Approaches to Language Typology 54), 63–86. Berlin: Mouton de Gruyter.

Nenadic, Oleg & Michael Greenacre. 2007. Correspondence analysis in R, with two- and three-dimensional graphics: The ca package. *Journal of Statistical Software* **20**(3), 1–13.

Nevis, Joel A. 1984. A non-endoclitic in Estonian. *Lingua* **64**(2–3), 209–224.

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

Pajusalu, Karl. 2007. Estonian dialects. In Mati Erelt (ed.), *Estonian Language*, 2nd edn. Tallinn: Estonian Academy Publishers.

Pajusalu, Renate. 2000. Indefinite determiners *mingi* and *üks* in Estonian. *Estonian: Typological studies IV*, 87–117.

Pajusalu, Renate. 2001. Definite and indefinite determiners in Estonian. In Enikő Németh (ed.), *Pragmatics in 2000: Selected Papers from the 7th International Pragmatics Conference*, vol. 2, 458–469.

Pajusalu, Renate. 2004. Viron üks ja kõik [Estonian one and all]. *Virittäjä* **108**(1), 2–23.

Pajusalu, Renate. 2009. Pronouns and reference in Estonian. *Sprachtypologie und Universalienforschung* **62**(1/2), 122–139.

Pajusalu, Renate. 2017a. Fraasid ja fraasiliikmed. Nimisõnafraas [Phrases and constituents: Noun phrases]. In Mati Erelt & Helle Metslang (eds.), 379–404.

Pajusalu, Renate. 2017b. Lause kommunikatiivne struktuur ja süntaktilised protsessid. Viiteseosed [Communicative sentence structure and syntactic processes: Reference links]. In Mati Erelt & Helle Metslang (eds.), 566–589.

Paldre, Leho. 1998. *Eitustundlikud üksused eesti keeles* [Negative polarity items in Estonian]. Unpublished MA thesis, University of Tartu.

Pant, Annika. 2018. Üle saja aasta õigekeelsusprobleeme: *gi-/ki*-liite paiknemine asesõnade käändevormides [Over a hundred years of orthographic problems: The location of the *gi-/ki-* suffix in the declension forms of pronouns]. *Oma Keel* [Your own language] **2**, 29–36.

Pant, Annika. 2020. *Pronoomenite keegi, miski, kumbki, ükski käändevormide kasutus tänapäeva eesti keeles* [The use of inflectional forms of the pronouns 'someone', 'something', 'either', 'none' in modern Estonian]. Unpublished BA thesis, University of Tartu.

Partee, Barbara H. 2008. Negation, intensionality, and aspect. In Susan Rothstein (ed.), *Theoretical and Crosslinguistic Approaches to the Semantics of Aspect* (Linguistics Today 110), 291–317. Amsterdam & Philadelphia: John Benjamins.

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

Pook, Hanna. 2021. Object case variation of the pronoun *mis* 'what' in spontaneous spoken Estonian and Estonian dialects. *Eesti ja soome-ugri keeleteaduse ajakiri* [Journal of Estonian and Finno-Ugric Linguistics] **12**(1), 259–301.

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

Ruutma, Mirjam, Aki-Juhani Kyröläinen, Maarja-Liisa Pilvik & Kristel Uiboaed. 2016. Ambipositsioonide morfosüntaktilise varieerumise kirjeldusi kvantitatiivsete profiilide abil [Descriptions of morphosyntactic variation in ambipositions using quantitative profiles]. *Keel ja Kirjandus* [Language and Literature] **2**, 92–113.

Saareste, Andrus. 1955. *Väike eesti murdeatlas* [Concise Estonian dialect atlas]. Uppsala: Almqvist & Wiksells.

Sammallahti, Pekka. 1977. Suomalaisten esihistorian kysymyksiä [Questions of Finnish prehistory]. *Virittäjä* **81**(2), 119–119.

Sang, Joel. 1983. *Eitus eesti keeles* [Negation in Estonian]. Tallinn: Valgus.

Seržant, Ilja A. 2015. The independent partitive as an Eastern Circum-Baltic isogloss. *Journal of Language Contact* **8**(2), 341–418.

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.

Tamm, Anne. 2015. Negation in Estonian. In Matti Miestamo, Anne Tamm & Beáta Wagner-Nagy (eds.), *Negation in Uralic Languages* (Typological Studies in Language 108), 399–432. Amsterdam: John Benjamins.

Tamm, Anne & Natalia Vaiss. 2019. *Setting the boundaries: Partitive verbs in Estonian verb classifications. Eesti Rakenduslingvistika Ühingu aastaraamat* [Yearbook of the Estonian Association of Applied Linguistics] **15**, 159–181.

Taremaa, Piia. 2017. *Attention meets Language: A Corpus Study on the Expression of Motion in Estonian*. Tartu University.

Uiboaed, Kristel. 2013. *Verbiühendid eesti murretes* [Verb constructions in Estonian dialects]. University of Tartu.

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.

Van Alsenoy, Lauren & Johan van der Auwera. 2015. Indefinite pronouns in Uralic languages. In Matti Miestamo, Anne Tamm & Beáta Wagner-Nagy (eds.), *Negation in Uralic Languages*, 519–546. John Benjamins.

Viikberg, Jüri. 2020. *Eesti murrete grammatika* [Grammar of Estonian dialects] (Eesti Keele Varamu VIII). Tartu: Tartu Ülikooli Kirjastus [University of Tartu Press].

Viitso, Tiit-Rein. 1985. Criteria for classifying dialects of Baltic Finnish languages. In *Dialectologia Uralica: Materials of the First International Symposium on the Dialectology of the Uralic Languages, 4–7 September 1984 in Hamburg*, 89–96.

Vitsberg, Eduard. 1958. *Eitus eesti murretes. Diplomitöö* [Negation in Estonian dialects: Diploma thesis]. Tartu: Tartu Riiklik Ülikool [Tartu State University].

Whaley, Lindsay J. 1996. *Introduction to Typology: The Unity and Diversity of Language*. Thousand Oaks: Sage Publications.

Zwarts, Frans. 1999. Polarity items. In Keith Brown & Jim Miller (eds.), *Concise Encyclopedia of Grammatical Categories*, 295–300. Oxford: Elsevier.

Cite this article: Pook H and Lindström L (2024). The use of the indefinite pronoun *keegi* 'someone' in Estonian dialects. *Nordic Journal of Linguistics* **47**, 192–223. https://doi.org/10.1017/S0332586522000221