With or without articles?  
A comparison of article-like determiners in Estonian and Finnish

HELEN HINT, TIINA NAHKOLA,  
RENA TE PAJUSALU  
University of Tartu

Abstract. In this paper, we compare the use and functions of definite and indefinite article-like determiners in Estonian and Finnish. Our main aim is to explore whether the factors that explain the choice of particular determiner forms are similar in Estonian and Finnish. We use a picture-sequence based elicitation experiment to collect spoken narratives from adult native speakers of Estonian and Finnish, and apply non-parametric tree and forest models to analyze the data. Our findings indicate that number of mention and animacy are important predictor variables in both languages, but their exact effect is divergent. We also find that in Finnish, case of the determiner NP proves to be an important factor, while in Estonian, syntactic role of the NP explains some aspects of determiner form choice. Nevertheless, the overall usage frequency of determiners is modest in the Estonian and Finnish data, and the process of grammaticalizing articles is only in initial stages in both languages.

Keywords: determiners; pronouns; grammaticalization; spoken narratives; Estonian; Finnish
1. Introduction

When we speak, one of our main needs is to refer to all kinds of entities around us. Various noun phrases (NPs) are used as referential devices, depending on the accessibility/givenness of the referent and the assumed cognitive states of the hearer (e.g., Ariel 1990; Gundel et al. 1993). An important concept related to reference is definiteness, which can be described as a grammatical system that signals the identifiability of the referent expressed with an NP in an ongoing discourse (Lyons 1999). By using a definite NP, the speaker expresses her assumption that the hearer can identify the referent based on the previous context; an indefinite NP is used when the speaker assumes that the referent cannot be identified by the speaker (Laury 2001a: 402–403).

The category of definiteness has grammatical realization in many languages that have specific morphemes – articles – for expressing the definiteness and/or indefiniteness of an NP. However, even in languages that lack grammatical articles, it is possible and common to express whether the referent is identifiable or not. Thus, it is necessary to distinguish grammatical and pragmatic definiteness. An NP is grammatically definite when definiteness is formally expressed by grammatical means; pragmatic definiteness, on the other hand, is indicated by the identifiability of the referent (Laury 2001b; Lyons 1999). Therefore, if the referent is identifiable to the hearer, the NP referring to this referent is also pragmatically definite; and if the referent has not yet been mentioned in the discourse and is not identifiable by other means, then the corresponding NP is pragmatically indefinite. In principle, NPs used to mention the referent for the first time are mostly indefinite, and all subsequent mentions are definite.

There are no grammatical articles in Estonian (est) or Finnish (fin) (e.g., Dryer 2013a, 2013b). However, linguistic devices for marking the (pragmatic) definiteness or indefiniteness of an NP still exist, e.g., case selection, word order, or determiners (see Chesterman 1991 for Finnish; Pajusalu 1997 for Estonian). For example, the use of different pronouns
as determiners in Estonian (e.g., üks ‘one’, see ‘this’) and in Finnish (e.g., se ‘it’, yks(i) ‘one’) has gained more attention recently (see Laury 1997; Pajusalu 2009). Determiners are a class of words that accompany NPs to indicate that the referent of an NP is identifiable. This class of determiners includes grammatical articles, demonstratives, quantifiers, and possessives. It is well known that certain determiners, especially demonstrative pronouns and the numeral ‘one’, often develop into grammatical articles through the process of grammaticalization (e.g., Heine & Kuteva 2006). This trend has been observed in Finnish (Laury 1997; Juvonen 2005) and also in Estonian (Pajusalu 1997; Pajusalu 2000). Yet, there are no studies analyzing the use of article-like determiners in parallel in present-day Estonian and Finnish, despite the fact that these languages are closely related and some of their determiners are very similar. The present paper aims to fill this gap.

This study has two major goals. Our first aim is to compare definite and indefinite determiners in Estonian and Finnish. We are interested in whether determiners are used to the same extent and in similar contexts in these languages. We also seek an answer to the question of whether the important predictor variables explaining the choice of particular determiner forms are similar or different in Estonian and Finnish.

More specifically, we compare the usage patterns of three forms.

i) Phonologically and functionally similar indefinite determiners, the numeral ‘one’ est üks/fiin yks(i).

ii) Phonologically similar, functionally slightly different demonstrative pronouns est see/fiin se functioning as definite determiners.

iii) Phonologically different, functionally similar possessives, est possessive pronoun oma/fiin possessive suffix. This is an interesting case, as the Finnish possessive suffix and Estonian pronoun oma differ in their morphologic realization which leads to different syntactic constructions and functions in an NP.

Our second intention is to implement a quasi-experimental methodology that permits a consistent and systematic comparison of languages.
By applying this method, we aspire to collect parallel and uniform language samples that can be used for studying all kinds of referential expressions in any language.

We begin by describing the background and giving an overview of the Estonian and Finnish determiners under investigation. The data collection and analysis methodology is described in Section 3. In Section 4, we present our results and illustrate them with examples from our data. Finally, a discussion and conclusions are offered in Section 5.

2. Background: Estonian and Finnish in the land of articles

European languages are typologically noteworthy due to the unusually wide use of articles. While the larger part of the world’s languages lack definite and/or indefinite articles, in Europe over 39 per cent of languages have both definite and indefinite articles and 15 per cent have only definite articles (Heine & Kuteva 2006: 98–99; see also Dryer 1989). Therefore, the existence of articles is an important feature of Standard Average European (Haspelmath 1998). However, Haspelmath (1998: 274) has claimed that in the eastern part of Europe, for example in the East Slavic, West Slavic and Finno-Ugric (except Hungarian) languages “there are no articles at all”. Yet, it is clear that language contact induced change and grammaticalization play an important role in the development of articles in eastern European languages, as well. The term grammaticalization marks the diachronic process of lexical items gradually changing into grammatical items, or one grammatical form into another (e.g., Heine et al. 1991).

Typological studies indicate that two wide-spread grammaticalization paths exist in the process of developing articles. Firstly, the numeral ‘one’ has grammaticalized into an indefinite article through being an indefinite pronoun in many languages, and has often preserved the same or a similar phonological form, e.g., German *ein*, French *un*, Turkish *bir*, English *a/one* (Givón 1981; Belaj & Matovac 2015; Lyons 1999: 95;
Weiss 2004). Secondly, definite articles are often historically developed from anaphoric adnominal demonstrative pronouns (Greenberg 1978; Diessel 1999; see also Heine & Kuteva 2006). Himmelmann (1996) also indicates that demonstrative expressions (including adnominal demonstratives) often have the function of tracking use which makes reference to major discourse participants, but only the presence of tracking use itself is not enough to say that a demonstrative has developed into an article.

The two quasi-universals (see Weiss 2004) of article grammaticalization also can be attested in Estonian and Finnish. It is true that Estonian and Finnish are traditionally described as languages without articles. However, the geographical location of Estonian and Finnish is a crucial aspect to bear in mind when speaking about article-like determiners in these two languages. While, according to Heine and Kuteva (2006: 110), the influence of articles in western European languages is the weakest on the northern part of eastern European languages, noticeable patterns regarding the use of article-like determiners in Estonian and Finnish cannot be neglected.

Importantly, previous studies describing the use of article-like determiners in Estonian and Finnish emphasize that both languages have historically been in steady continuous contact with European article-languages: Estonian was under strong German influence until the late 19th century, Finnish has existed in durable contact with Swedish1 (Nordlund et al. 2013). In German and Swedish, both indefinite and definite articles are present. Furthermore, due to the prevalent use of English in the contemporary world, Estonian and Finnish are once again in the sphere of an article-language.

It has been claimed that the Estonian numeral and indefinite determiner üks ‘one’ has the potential for developing into an indefinite article (Pajusalu 2000; Pajusalu 2009). Similarly, researchers have described the Finnish numeral and indefinite determiner yks(i) ‘one’ as a typical candidate for becoming an indefinite article in spoken contemporary language

1 In addition to Finnish, Swedish is also an official language of Finland.
(Vilkuna 1992; Juvonen 2005). While yks(i) is not that frequent according to the Finnish reference grammar (VISK 2008: § 1418), it nevertheless occurs in contexts very typical of an indefinite article. More precisely, according to Vilkuna (1992: 32), yks(i) is speaker-definite but hearer-indefinite, i.e., the speaker signals ‘I know the referent, but you don’t.’ However, the use of an indefinite determiner depends mostly on speech situation and context; indefinite determiners are more common in spoken (colloquial) language and it is also not obligatory to mark the indefiniteness of an NP. In addition, there are other indefinite determiners available, e.g., mingi ‘some’ in Estonian, which is mostly used in informal contexts (Pajusalu 2000; Pajusalu 2009), and eräs ‘one’ in Finnish, which is more formal in style than yks(i) (Vilkuna 1992: 32; VISK 2008: §749).

While the usage patterns of Estonian mingi NPs and Finnish eräs NPs are more specific and are not discussed further in this paper, it should be noted that the usage frequency of mingi and eräs as determiners in our Estonian and Finnish data, respectively, indicates that our elicited narratives fall somewhere between formal and informal language use.

Definite article-like determiners are also phonetically similar in Estonian and Finnish. Standard Estonian uses the proximal demonstrative see ‘this’ adnominally as a marker of definiteness in contexts where the referent of a seeNP has been previously mentioned in preceding discourse (Pajusalu 1997; Pajusalu 2009). However, the definite article has not yet fully grammaticalized in Estonian (Pajusalu 1997: 173). Regarding Finnish, Ritva Laury (1997), following Greenberg’s (1978) typology, has suggested that the anaphoric demonstrative se ‘it’ has already grammaticalized into a stage 1 definite article in spoken language. Yet, there are also different opinions about Finnish seNPs which describe se instead as a demonstrative that can, among other functions, be used as a definite determiner (Juvonen 2000; Larjavaara 2001). The Finnish reference grammar (VISK 2008: § 1418) does not state whether the determiners se (and yksi) are already grammaticalized into articles, but according to frequency in spoken Finnish, se is more article-like than yks(i).

---

2 We would like to thank an anonymous reviewer for this valuable comment.
While no studies suggest a fully grammaticalized category of articles in Estonian or Finnish, there is no doubt that article-like determiners are an integral part of the Estonian and Finnish systems of (in)definite-ness markers. Nevertheless, the current stage of the grammaticalization of Estonian and Finnish determiners is debatable. Heine and Kuteva (2006: 119) have placed both languages in the incipient category regarding the grammaticalization of the definite article. Likewise, they treat Estonian and Finnish as undergoing an incipient grammaticalization of indefinite articles. However, they admit that their data are too scarce for making any far-reaching conclusions about whether Estonian and Finnish article-like determiner uses are major usage patterns or not (Heine & Kuteva 2006: 133).

A crucial point to be kept in mind is that although Estonian and Finnish are genetically and typologically similar and have emerged from the same Proto-Finnic language, the pronominal systems of Estonian and Finnish show important differences. According to the traditional view, Standard Estonian has two demonstrative pronouns: proximal see 'this' and distal too 'that', which are used similarly in anaphoric function (EKG II 1993: 209). But in contemporary Estonian the uses of demonstrative too are rather restricted and proximal see is the only demonstrative in many varieties of Estonian (Pajusalu 2006; Pajusalu 2009; see also Reile 2015). While distal too can be used adnominally as a determiner, this kind of use is rather rare and instances of the tooNP have retained the deictic restriction to far entities in the case of contrasted referents.

The Standard Finnish demonstrative system is comprised of three demonstrative pronouns: proximal tämä (colloquial tää) 'this', anaphoric se 'it/this', and distal tuo (colloquial toī) 'that'. As of now, researchers have proposed interactional explanations for usage of Finnish demonstratives instead of a spatial distance-based explanation (Laury 1997; Seppänen 1998; Etelämäki 2006; Priiki 2017). It is stated that tämä is used when the referent is in the sphere of the speaker, tuo places the referent outside of the speaker's sphere, and se is used for referents in the addressee's sphere (Laury 1997: 59). Other researchers say that tämä is mostly used for
referents which are identifiable on the basis of the utterance they belong to, while *tuo* is identifiable in a larger context (Etelämäki 2006; Priikki 2017). While some researchers claim that there are two definite determiners, *se* and *tämä*, that compete for the status of the definite article in Finnish (Juvonen 2000), others see adnominal *tämä* instead as a marker of prominence and *se* as a definite article (Laury 1997).

To the best of our knowledge there are no contrastive studies that compare Estonian and Finnish determiners in contemporary (spoken) language in similar texts/language situations. The present study aims to fill this gap, by giving an up-to-date comparison of the system of article-like determiners in these two cognate languages. There is, however, a previous contrastive study concerning determiners in old Estonian and Finnish literary texts (Nordlund et al. 2013). This study concentrates on written texts from the 18th century which, as the authors point out, are often written by non-native speakers and are often translations from German (in the case of Estonian), Swedish (in the case of Finnish) or from Latin and Greek. This study shows that the use of determiners was already systematic in the 18th century Finnish and Estonian, and while the use of determiners is largely borrowed from article-languages, there are also contexts of “native” use which is not directly copied from source texts (Nordlund et al. 2013).

### 3. Method

We used a picture-sequence based narrative elicitation experiment to collect spoken Estonian and Finnish narratives in order to scrutinize different referential expressions, including NPs with article-like determiners.\(^\text{3}\) This particular methodology is adapted from Koster et al. (2011) who used a narrative production experiment for studying children’s acquisition of subject pronouns. A classic study using a similar

\(^\text{3}\) The experiment was originally designed for studying all types of referential NPs, including fullNPs, pronouns, demonstratives, etc. For the present study, only NPs with (or without) a determiner were analyzed.
method was “The Pear Stories” project by Chafe et al. (1980), in which a short film was used to collect narratives from a number of languages to study the variation of simple stories across languages. Experimental elicitation was chosen for the purpose of getting a well-structured dataset, which is also well-suited for comparison among languages. In a contrastive study, using a uniform method allows one to examine languages more rigidly, since the context, referents, and the purpose of narratives remain similar throughout the dataset.

3.1. Participants

40 adults volunteered for participation in the study: 20 native Estonian speakers (13 females, 7 males) and 20 native speakers of Finnish (13 females, 7 males). The Estonian speakers were 25–47 years old (mean age 32), the Finnish speakers were 20–80 years old (mean age 46).

3.2. The procedure

Each participant was shown three different storybooks with 6 pictures in each book and 1 picture per page. The stories are called the apple story (AS), the bike story (BS), and the kite story (KS) (the pictures are presented in the Appendix). Participants were asked to tell a short story based on each book. To avoid text-external reference (e.g., body language, pointing to the pictures, etc.), participants were requested to speak so that a person who is not present at the moment would understand the story if she listened to the recording later. Each participant was tested individually in a quiet room and the test sessions were audio-recorded. The experimental sessions took approximately 10–15 minutes.

All picture sequences shared the same internal organization. In the first and second pictures, there was always one character (boy1)

4 This condition might have triggered the use of a more formal register than in everyday conversation. We thank an anonymous reviewer for pointing this out.
performing an action. The second character (boy2) was introduced in the third picture and in the fourth and the fifth pictures this second character was the actor, while the first character remained passive in the background. Then the first character (boy1) was shown alone in the final picture again. Both characters were of the same gender so that in the future these pictures can also be used for studying and comparing languages which do have a grammatical gender distinction.

Two animate referents, boy1 and boy2, were the same in each storybook; they are “competitive” referents, since both take the prominent position in some part of the story. Additionally, we included three inanimate referents from each book: kite, big tree, and stick from the kite story; watering jug, apple tree, and apple from the apple story; and bike, broken front wheel, and new front wheel from the bike story. The inanimate referents are chosen based on their importance to the storyline; we expected that these referents would be mentioned more than once in the stories, forming reference chains in which their accessibility and corresponding referential expression change. Also, these inanimates can take either the subject or object position in each story.

### 3.3. Coding the variables

The recordings were transcribed and coded by Estonian and Finnish native speakers, accordingly. From the Estonian sample, we had to exclude 1 narrative, and from the Finnish sample 6 narratives, due to unconventional strategies of using referential devices (using only proper names, using direct speech and first and second person, or combining all three stories into one connected discourse). Therefore, for the final analysis we had 59 Estonian and 54 Finnish narratives.

For the purpose of the present study, we conducted the analysis with only those NP types where the presence or absence of the determiner is meaningful, i.e., we were only interested in NPs that can (in principle) syntactically occur with a determiner. Therefore, we analyzed

i) bare NPs (bareNP; e.g., est poiss, fin poika, ‘(the) boy’); ii) NPs with
an adjective attribute (adjNP; e.g., EST väike poiss, FIN pieni poika, ‘(a/the) small boy’); iii) NPs with a nominal genitive attribute (genNP; e.g., EST poisi ratas, FIN pojan polkupyörä, ‘(the) boy’s bike’); iv), and NPs with a determiner (detNP; e.g., EST see poiss, FIN se poika, ‘the/this boy’). Personal pronouns, bare demonstrative pronouns, bare demonstrative adverbs, some NPs with a genitive attribute (see below), proper names, relative clauses and zero reference were not included, since these do not occur with (article-like) determiners.\footnote{There actually are instances of proper names with determiners in Estonian and Finnish (e.g., EST see Peeter, FIN se Petteri ‘that Peter’), but as this kind of use marks demonstrative reference to a specific mention of the name within the text, rather than definiteness in particular (Pajusalu 1997: 153), such uses are not included in the present analysis.}

We excluded NPs with a pronominal genitive attribute (e.g., EST tema ratas, FIN hänen pyörä(nsä) ‘his bike’) and NPs with a proper name as a genitive attribute (EST Peetri lohe, FIN Pekan leija ‘Pete’s kite’) from the analysis. Also, as our data included only four NPs with an attribute that is located after the noun it is modifying, and as such NPs were only present in the Estonian data (EST poiss kastekannuga ‘a/the boy with a watering can’), we decided to code these NPs as NPs with an adjective attribute.

These selection criteria left us 676 Estonian NPs and 680 Finnish NPs for the final analysis. The distribution of different NP types included in the analysis is presented in Table 1.

**Table 1. Distribution of analyzed phrase types in the data**

<table>
<thead>
<tr>
<th>Language</th>
<th>adjNP</th>
<th>detNP</th>
<th>genNP</th>
<th>bareNP</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EST</strong></td>
<td>104 (15,4%)</td>
<td>188 (27,8%)</td>
<td>10 (1,5%)</td>
<td>374 (55,3%)</td>
<td>676 (100%)</td>
</tr>
<tr>
<td><strong>FIN</strong></td>
<td>119 (17,5%)</td>
<td>121 (17,8%)</td>
<td>19 (2,8%)</td>
<td>421 (61,9%)</td>
<td>680 (100%)</td>
</tr>
</tbody>
</table>

For the further statistical analysis, all NPs referring to the abovementioned important referents in the storyline were first coded and tagged.
for the following predictor variables: animacy, other animate entities in the utterance, number of mention, and referential distance to the previous mention (pragmatic and semantic factors); number, case of the NP, syntactic role, and clause type (grammatical factors); and participant’s age, gender, and education (social factors). Table 2 summarizes the pragmatic/semantic and grammatical predictor variables and their levels.

Table 2. Predictor variables used in this study

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Variable levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animacy</td>
<td>Animacy of the referent</td>
<td>anim (animate), inanim (inanimate)</td>
</tr>
<tr>
<td>Other Anim</td>
<td>The presence of other animate referents in the clause</td>
<td>yes, no, inanim (only inanimate referents)</td>
</tr>
<tr>
<td>Mention No</td>
<td>Number of mention in a row of the reference chain for the particular referent</td>
<td>1, 2, 3, further (all further mentions), impl (previously implicitly mentioned), other</td>
</tr>
<tr>
<td>RefDist</td>
<td>Referential distance</td>
<td>0 (same utterance), 1 (preceding utterance), 2 (2 utterances back), 3 (3 utterances back), further (4 or more utterances back), no (no previous mentions)</td>
</tr>
<tr>
<td>No</td>
<td>Number</td>
<td>sg (singular), pl (plural)</td>
</tr>
<tr>
<td>Case</td>
<td>Case of the NP</td>
<td>nom (nominative), gen (genitive), part (partitive), ill (illative), ine (inessive), ela (elative), all (allative), ade (adessive), abl (ablative), ess (essive), com (comitative)</td>
</tr>
<tr>
<td>SyntRole</td>
<td>Syntactic role</td>
<td>subj (subject), obj (object), genAtr (genitive attribute), other</td>
</tr>
<tr>
<td>Clause Type</td>
<td>Clause type of the utterance containing the NP</td>
<td>main (main clause), sub (subordinate clause)</td>
</tr>
</tbody>
</table>

*The Estonian case category has 14 members, while Finnish has 14 or 15 members; however, due to the nature of the production task used, not all cases were present in the data regarding the crucial referents.*
Among the grammatical factors, separating case and syntactic role is somewhat redundant regarding nominative, genitive and partitive cases, as subjects are prototypically in the nominative case, and objects in the genitive or partitive case in Estonian and Finnish. Yet, semantic cases, especially locatives, expressing other syntactic roles are expected to show greater variance in this respect, and this also may be reflected in the choice of determiners.

Pragmatic and semantic factors, such as animacy, number of mention, and referential distance, have proven relevant in the choice of referential devices (see e.g., Givón 1983; Gundel et al. 1993; Siewierska 2004). For languages without a grammatical gender distinction, it is sometimes also important to take into account whether there are one or several animate entities in an utterance. In Estonian, for example, it has been shown that when there is more than one animate entity in an utterance, a personal pronoun usually refers to the most prominent character (i.e. protagonist), which is usually a subject, while a demonstrative pronoun typically refers to a less prominent minor character (Pajusalu 2009). Similarly, in Standard Finnish the demonstrative tämä can refer to non-topical referents in the background (Varteva 1998). Therefore, differences might also emerge when considering the use of determiners.

3.4. Data analysis method

To predict the most important factors and associations between multiple possible factors affecting the presence/absence and type of determiners in Estonian and Finnish, we applied two explorative classification techniques: conditional inference recursive partitioning tree models (Hothorn et al. 2006) and conditional random forest analysis (Breiman 2001; Strobl et al. 2008; see also Strobl et al. 2009). We were also interested in determining whether the significant factors influencing the choice of determiners are similar or different in Estonian and Finnish.

Conditional inference trees and random forests as non-parametric methods for handling complex data with categorical variables (Strobl
et al. 2009; Tagliamonte & Baayen 2012) were well-suited for our data. Recursive partitioning of conditional inference trees is implemented by recursive binary splitting of the data based on significance testing (Hothorn et al. 2006). The analysis results in a dendrogram that captures the interactions between predictor variables in a visually clear and straightforwardly interpretable manner (Strobl et al. 2009). However, as the simple tree models are relatively unstable and show high variability, the complementary information provided by the random forests method is needed (Ibid.). The random forest method constructs a larger set of trees and, based on the average of all the trees, selects the important variables that best classify the data; the relative variable importance then can be calculated (Breiman 2001; Tagliamonte & Baayen 2012). These methods, also called “tree & forest” (see Janda 2013: 26), have been successfully implemented in linguistic studies as well (e.g., Tagliamonte & Baayen 2012; Baayen et al. 2013; Lindström & Vihman 2017; Taremaa 2017; Priiki 2017).

The open source statistical environment R (R Core Team 2016) package `party` was used for the computations, the `ctree` function was used for conditional inference trees, and `cforest` was used for random forest analysis. Variable importances were calculated using the `varimp` function.

4. Results

In this section, we describe the results of the quantitative analysis of our narrative production task and provide examples of different determiners present in our data.

4.1. Results of Estonian and Finnish tree & forest analysis

The focus of the present study was to observe the usage patterns of NPs with different determiners compared to NPs without determiners (øNP). Thus, we entered the determiner form (DetForm) as the dependent variable with four levels in our models. The levels were i) øNP, ii)
seeNP, iii) üksNP, and iv) otherNP in Estonian, and i) øNP, ii) seNP, iii) tämäNP, and iv) otherNP in Finnish. We coded the two most frequent determiners in both languages as separate levels and grouped together all less frequent determiners as ‘others’.

The distribution of determiner forms in the data is presented in Table 3. As can be seen from this table, NPs with determiners are less common than øNPs in both Estonian and Finnish. However, out of all NPs that are syntactically compatible with a determiner, 27.7% in Estonian and 21.3% in Finnish are actually used with one, which suggests that determiners are not just occasional but have a specific grammatical or discourse-related purpose. It should be noted that there were only 7 occurrences of the Finnish indefinite determiner yks(i) in our data. Consequently, these are coded as otherNPs for the present analysis (see Table 5). Note also that in some contexts of language use, e.g., in informal spoken narratives, the Estonian demonstrative see ‘this’ corresponds to both se ‘it’ and tämä ‘this’ (and sometimes even tuo ‘that’) in the Finnish system of demonstratives, which makes the comparison of determiners more complex.

**Table 3. Distribution of NPs with and without determiners in the data**

<table>
<thead>
<tr>
<th>Language</th>
<th>øNP</th>
<th>otherNP</th>
<th>seeNP (EST)/seNP (FIN)</th>
<th>tämäNP (FIN)</th>
<th>üksNP (EST)/yks(i)NP (FIN)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>EST</td>
<td>489</td>
<td>50</td>
<td>94</td>
<td>x</td>
<td>43 (6,4%)</td>
<td>676</td>
</tr>
<tr>
<td></td>
<td>(72,3%)</td>
<td>(7,4%)</td>
<td>(13,9%)</td>
<td></td>
<td></td>
<td>(100%)</td>
</tr>
<tr>
<td>FIN</td>
<td>535</td>
<td>59</td>
<td>42</td>
<td>37</td>
<td>7 (1%)</td>
<td>680</td>
</tr>
<tr>
<td></td>
<td>(78,7%)</td>
<td>(8,7%)</td>
<td>(6,2%)</td>
<td>(5,4%)</td>
<td></td>
<td>(100%)</td>
</tr>
</tbody>
</table>

Table 3 indicates that aside from the most frequent major determiners see/se, fin tämä and est üks, the group consisting of other determiners is substantial. Yet, the exact content of this group is diverse and it combines determiners with very different meanings and discourse functions. It is beyond the scope of this paper to explain the role of all determiners.
in this group, but as an overview, Table 4 and Table 5 summarize other determiners found in Estonian and Finnish, respectively.

**Table 4. Other determiners in Estonian**

<table>
<thead>
<tr>
<th>Est determiner</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>enda ‘(his/her) own’</td>
<td>1</td>
</tr>
<tr>
<td>keegi ‘someone’</td>
<td>1</td>
</tr>
<tr>
<td>mingi ‘some’</td>
<td>8</td>
</tr>
<tr>
<td>mingisugune ‘a kind of’</td>
<td>1</td>
</tr>
<tr>
<td>oma poss.pron ‘one’s’</td>
<td>29</td>
</tr>
<tr>
<td>seal local pro-adverb ‘there’</td>
<td>1</td>
</tr>
<tr>
<td>sealt local pro-adverb ‘(from) there’</td>
<td>2</td>
</tr>
<tr>
<td>seesama ‘the same’</td>
<td>1</td>
</tr>
<tr>
<td>sina local pro-adverb ‘(to) there’</td>
<td>5</td>
</tr>
<tr>
<td>too ‘that’</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
</tr>
</tbody>
</table>

**Table 5. Other determiners in Finnish**

<table>
<thead>
<tr>
<th>Fin determiner</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>eräs ‘one’</td>
<td>1</td>
</tr>
<tr>
<td>possessive suffix</td>
<td>25</td>
</tr>
<tr>
<td>semmonen ‘that kind of; such’</td>
<td>2</td>
</tr>
<tr>
<td>siellä ‘(in) there’</td>
<td>2</td>
</tr>
<tr>
<td>sieltä ‘from there’</td>
<td>13</td>
</tr>
<tr>
<td>tommonen ‘that kind of; such’</td>
<td>4</td>
</tr>
<tr>
<td>tuo ‘that’</td>
<td>1</td>
</tr>
<tr>
<td>tuollainen ‘that kind of; such’</td>
<td>2</td>
</tr>
<tr>
<td>tällainen ‘this kind of’</td>
<td>1</td>
</tr>
<tr>
<td>tämmönen ‘this kind of’</td>
<td>6</td>
</tr>
<tr>
<td>tämä sama ‘this same’</td>
<td>1</td>
</tr>
<tr>
<td>täältä ‘from here’</td>
<td>1</td>
</tr>
<tr>
<td>yks(i) ‘one’</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
</tr>
</tbody>
</table>
We were interested in the factors which best explain the choice of determiner forms in Estonian and in Finnish, and whether the important factors coincide or show considerable differences in these languages. The following analysis explains these factors in more detail. The recursive partitioning tree and random forest analysis models were identical for Estonian and Finnish, including all predictors described in Section 3.3. As the tree models exclude factors with no significant effect from the graph, the following tree figures present only statistically significant factors in choosing the particular referential form.

4.1.1. Estonian results

The conditional inference tree of the Estonian data presented in Figure 1 shows that in Estonian, the first partition splitting the data into two subsets is made by number of mention (Node 1). The importance of number of mention as the main factor is adaptable to the notion of definiteness: entities mentioned for the first time are mostly indefinite and subsequent mentions are always definite. Therefore, the motivation behind this split is the presence of definite seeNPs and indefinite üksNPs as the most frequent NPs with determiners in the Estonian data.

When the referent is mentioned for the first time (i.e., as indefinite or implicitly known), animacy is the next most important factor (Node 2). In the case of animate referents, the probability of using an üksNP is relatively high, whereas NPs with other determiners are infrequent and, as one would intuitively expect, definite seeNPs do not occur at all (Node 3). When the referent is inanimate, the probability of using (indefinite) determiners for the first mention is lower compared to animate referents (Node 4).

In the right-hand branch of subsequent mentions (second and further), the split is made by syntactic role (Node 5). Subjects and genitive attributes are first separated from objects and other syntactic roles; Node 9 indicates that when an NP functions as a subject or genitive attribute
in an utterance, and if it happens to be a detNP, then the determiner will be see. As there were only a few NPs with genitive attributes in our data (see Table 1), this node mostly characterizes subjects.

Another split by syntactic roles is made for separating objects from other syntactic roles (Node 6). With other syntactic roles, different determiners can be used and no prominent determiner emerges in this node (Node 7). Objects, on the other hand, are relatively often expressed with seeNPs (Node 8), although other determiners are also possible. Objects are even more likely to be referred to with a seeNP than subjects, and the overall proportion of detNPs compared to øNPs is the highest among subsequent mentions in Node 8. As anticipated, üksNPs do not occur with subsequent (definite) mentions.

However, it is important to keep in mind that we were focused on the use of detNPs compared to NPs without determiners. As Figure 1 reveals, NPs without a determiner are by far the most frequent referential devices in every terminal node. Therefore, in our data there are no contexts where the indefinite determiner üks or the definite determiner see function as obligatory (grammatical) markers of (in)definiteness.

We also conducted a random forest analysis in order to support the results obtained from the conditional inference tree, and the variance importance graph retrieved is presented in Figure 2. Predictors depicted to the right of the vertical line are significant and significance increases when one moves upwards and to the right (i.e. farther from zero). The predictors around zero remain unimportant in explaining the choice of determiner forms. Thus, Figure 2 suggests that age, number of mention, and syntactic role are the three most influential predictors of determiner form in Estonian. The importance of a socio-linguistic factor age is intriguing. The data indicate that most determiners are used by participants 27–34 years old. However, the present data are too limited to make more far-reaching conclusions regarding how age affects the use of determiners. Referential distance, which did not appear as an important predictor on the conditional inference tree, also affects the use of determiners. This is further illustrated by the Estonian examples in section 4.2.
Figure 1. Conditional inference tree for the distribution of øNPs and detNPs in Estonian data: DetForm ~ Age + Gender + Education + Animacy + OtherAnim + MentionNo + RefDist + Case + Num + SyntRole + ClauseType

Figure 2. Conditional variance importance in predicting øNPs and detNPs for the random forest of Estonian data. Predictors to the right of the vertical line are significant
To summarize the Estonian results, the three most important linguistic factors affecting the choice of determiner are number of mention, syntactic role, and animacy. Animacy is only important among first mention NPs and syntactic role in case of subsequent mentions.

i) The NP with an indefinite determiner üks most probably refers to an animate referent mentioned for the first time.

ii) The definite determiner see is most characteristic of subsequent-mention NPs in subject or object position.

iii) Indefinite inanimate referents and definite NPs in non-subject and non-object position do not show strong preferences for particular determiners, and the probability of using a determiner with these referents is lower.

4.1.2. Finnish results

The results of the Finnish tree & forest analysis also highlight the importance of number of mention and animacy as the predictor variables, as in Estonian. Yet, unlike in Estonian, case is deemed more important in Finnish than syntactic role. Moreover, the exact effect of predictor variables is slightly different.

The Finnish conditional inference tree, presented in Figure 3, shows that the first predictor in the choice of Finnish determiner forms is case (Node 1). Two cases, ablative and elative, are split into one distinct group and, according to the tree, NPs in these cases occur more probably with other determiners, including adverbial determiners sieltä and tältä, while demonstrative determiners showing case agreement, e.g. siltä (se.ABL) and tältä (tämä.ABL) are marginal (Node 2). Remarkably, this terminal node suggests the highest probability of using a determiner in Finnish. The underlying reason here is that this group assembles NPs with the adverbial determiners, mostly sieltä. Furthermore, upon examining the data more closely, we found that this node covers predominantly inanimate referents: only 4 NPs referred to animate entities in this node. The reason is that with animates, the case-agreeing forms siltä or
tältä would probably be more common than sieltä or täältä, which are locative (spatial) in meaning.\footnote{We thank an anonymous reviewer for this remark.}

Within the group of other cases (right branch), the next split is made by number of mention (Node 3). As in Estonian, a contrast exists between first and subsequent mentions; however, in Finnish, implicitly known first-mention referents belong together with subsequent, non-first-mentions. When the referent is mentioned for the first time, other determiners are preferred (Node 9). Again, the effect of definite and indefinite reference is meaningful, since the definite determiners se and tämä are not typical for first-mention (indefinite) referents.

Subsequent mentions are additionally influenced by the effect of animacy (Node 4): tämä is the preferred determiner for animate referents (Node 5) and se is slightly more frequent in referring to inanimate referents (Nodes 7 and 8). A similar effect for animacy was also present in Juvonen’s (2000) study concerning Finnish definite determiners. While inanimate referents are further split by case (Node 6), se is still the predominant determiner in both terminal nodes. Interestingly, tämäNPs are not used when the NP is in allative, illative, inessive, or nominative case (Node 7). Nevertheless, terminal nodes 5 and 8 have rather minor variation in determiner form choice and therefore it is difficult to decide the real difference between seNPs and tämäNPs in Finnish based on these data.

Finnish results do resemble the outcome of Estonian in that the majority of referential NPs are used without any determiners: in every terminal node of the conditional inference tree, øNPs are dominant.

A forest analysis was also implemented for Finnish and the conditional permutation-based variance importance graph is displayed in Figure 4. Similar to Estonian, age is displayed as the most important predictor of determiner form choice. While the qualitative inspection of the data suggests that participants older than age 50 tend to use fewer determiners, conclusive inferences concerning the effect of age are beyond the scope of the present study.
Figure 3. Conditional inference tree for the distribution of ØNPs and detNPs in Finnish data: DetForm ~ Age + Gender + Education + Animacy + OtherAnim + MentionNo + RefDist + Case + Num + SyntRole + ClauseType

Figure 4. Conditional variance importance in predicting ØNPs and detNPs for the random forest of Finnish data. Predictors to the right of the vertical line are significant.
Language-related factors, case and number of mention, also appear as significant, concurring with the conditional inference tree in Figure 3. The importance of other variables which position around zero is considered insignificant.

To recap, case and number of mention are the most influential linguistic factors distinguishing Finnish determiners, while animacy shows particular effect in the Finnish conditional inference tree.

i) If an NP in the allative or elative case (i.e., non-subject, non-object) occurs with a determiner, then this determiner is probably an adverbial (e.g., sieltä).

ii) Referents just introduced into discourse (i.e., first mentions) prefer other determiners (e.g., yksi, tämmönen etc.).

iii) Subsequent mentions (i.e., definite NPs) do not show very strong preferences in determiner form choice; however, animate referents slightly favor tämäNPs and inanimate referents favor seNPs.

4.2. Examples of determiner use in Estonian and Finnish

In this section, we present excerpts from our Estonian and Finnish narratives to illustrate determiner use in both languages.

4.2.1. Estonian see

According to the previous analysis, the use of Estonian seeNPs is more common with previously mentioned subject and object referents.

In Example (1), two subsequent utterances exhibit seeNPs: the subject referent is mentioned with a seeNP in both utterances. Note that the subjects are different in these utterances. Both referents are accessible and definite, and as topical characters in the story, they are mostly referred to by using personal pronouns in the preceding text. However, utilizing pronouns in context where two animate referents are present at the same time would probably confuse the hearer, so the speaker chooses
full NPs here. By adding a determiner *see*, the speaker indicates that the referents are already known to the hearer and that they are highly accessible.

Remarkably, the Estonian conditional random forest analysis (see Figure 2) indicated that referential distance is also an important predictor in the choice of determiners. Example (1) illustrates this result: the referent (*boy* A) has not been mentioned in several preceding utterances, and it is introduced again by using a *see* NP.

(1) EST8, AS

\[
\begin{array}{llllllllll}
\text{sis} & \text{ta} & \text{võtab} & \text{selle,} & \text{se} & \text{nokamütsiga poiss} \\
\text{then} & \text{3SG.NOM} & \text{take.3SG.PRS} & \text{this.NOM} & \text{see.NOM} & \text{cap.COM} & \text{boy.NOM} \\
\text{võtab} & \text{oo} & \text{õunapuu} & \text{küljest} & \text{ühe} & \text{õuna}.
\end{array}
\]

\[
\begin{array}{llllllllll}
\text{take.3SG.PRS} & \text{umm} & \text{apple.tree GEN} & \text{from} & \text{there} & \text{one.NOM} & \text{apple.GEN} \\
\text{see} & \text{kastekannuga} & \text{mees} & \text{just} & \text{vaatab} & \text{et} \\
\text{see.NOM} & \text{watering can.COM} & \text{man.NOM} & \text{right} & \text{look.3SG.PRS} & \text{that} \\
\text{vo vo vo} & \text{mis} & \text{sa} & \text{teed}^9 \\
\text{PRCL} & \text{what} & \text{2SG.NOM} & \text{do.2SG.PRS}
\end{array}
\]

‘then he takes this, the boy with a cap takes umm an apple from the apple tree. the man with a watering jug just looks that wo wo wo what are you doing’

---

^ EST – Estonian, FIN – Finnish; number indicates the participant; AS – apple story, BS – bike story, KS – kite story. From all the recorded narratives, a detailed transcriptions are constructed according to the conventions of speech. However, for the reasons of clarity and readability, pauses and other prosodical phenomena are not included in presenting the examples here, since in the analysis we only accounted for the presence/absence of a determiner in an NP. In the glossings, Estonian determiner *see* and Finnish determiners *se* and *tämä* are left untranslated, since in many contexts there are more than one possible translations to a particular form (i.e., ‘the’, ‘that’, or ‘it’).

^ In this example, the detNP *se nokamütsiga poiss* occurs in a position of a delayed theme; however in this study we do not analyse the specific syntactic environments of detNPs, but only the referential properties of detNPs.
Definite objects are also commonly referred to with a seeNP. Example (2) shows that the referent in object position (kite) even is referred to with a seeNP twice. The kite has already been introduced into the narrative at the beginning of the narration. As there is a minor referential distance (the last mention of the kite is in a subordinate clause, a main clause separates the two mentions), the speaker has selected a seeNP instead of a demonstrative pronoun.

(2) EST15, KS

ning aitas ä see.selle see.lohe puu otsast
and help.3SG.PST umm see.gen kite.gen tree.gen from upon
alla ning ää see.lohe see.manik oli önnelik et
down and umm kite owner.nom be.3SG.PST happy that
ta see.selle see.lohe kätte sai
3SG.NOM see.gen kite.gen hand.ill get.3SG.PST

‘and helped umm the kite down from the tree and umm the kite owner was happy that he got the kite back’

For other syntactic roles, such as an adverbial expressing an instrument meaning in Example (3), it seems that the use of the seeNP is an instance of the tracking use of a demonstrative (see Himmelmann 1996). The previous mention of the same referent (watering can) is made with an øNP, although this referent has previously been mentioned and is fully identifiable to the hearer. So, the development of demonstratives into articles is not supported by this example.

(3) EST13, AS

ta läheb kraani juurde jaa täidab kastekannu.
3SG go.3SG.PRS faucet.gen to and fill.3SG.PRS watering can.gen
jaa see.selle see.kastekannu see.kastab toda önnapuud.
and see.gen watering can.com water.3SG.PRS that.part apple.tree.part

‘(he) goes to the faucet and fills the watering can and waters this apple-tree with the watering can’
4.2.2. Finnish *se*

Regarding Finnish, the analysis suggested that *se*NPs refer more often to inanimate rather than animate referents. At the same time, the determiner *se* also can occur with animate referents. In Example (4), the kite is referred to with a *se*NP. The preceding mention of the kite was made five utterances earlier with an anaphoric pronoun *se*. Now the same referent is brought back to attention and a definite determiner is necessary in order to indicate that the same kite has been already mentioned.

Interestingly, the other determiner *tämä* is also present in the same utterance, referring to an inanimate entity (*stick*), as well. Unlike the kite, the stick has been just mentioned for the first time in the preceding utterance. Here, the difference can be viewed as a distinction between open and closed reference (Etelämäki 2006): while the kite is already known from the previous discourse and does not require additional explanation (i.e., closed reference), the stick has just been introduced to the text and might need further definition (i.e., open reference).

(4) FIN3, KS

```
ja kaverilta löytyy sit tämmönen pitkä tikku
and guy.ABL find.3SG.PST then this kind of.NOM long.NOM stick.NOM

ja sitte kaveri tällä tikulla auttaa sen leijan
and then guy.NOM stick.ADE help.3SG.PRS gen kiteGEN

pois sieltä puusta
away from there tree.ELA
```

‘and then the guy has this kind of a long stick and then the guy helps the kite down from the tree with/that stick’

Another instance of a *se*NP referring to inanimate referent is found in (5). Unlike the *se*NP in the previous example, here a definite determiner is used already for the second, not further mention. Second mentions are often made using detNPs. Additionally, a one-utterance referential distance from a preceding mention also gives rise to a need for a definite determiner.
and give.3SG.PRS totally watering can.ADE watering boy.ALL.PRTCL.PRTCL
omenan sieltä puusta ja tarina loppuu taas
apple.GEN from there tree.ELA and story.NOM end.3SG.PRS again
onnellisesti, poika pääsee syömään sitä omenaa.
happily boy.NOM get.3SG.PRS eat.INFILL se.PART apple.PART
‘and gives to this boy with a watering can umm an apple from the tree and
the story ends happily again, the boy gets to eat the apple’

Nevertheless, seNPs are also perfectly suitable for use with animate enti-
ties, as can be seen in (6).

(6) FIN 3, BS
sit siihen tulee toinen kaveri, tähän pojan
then to there come.3SG.PRS another.NOM guy.NOM to here boy.GEN
luo, ja sitte poika valittaa sille kaverille
at and then boy.NOM complain.3SG.PRS se.ALL guy.ALL
että pyörä meni rikki
that bike.NOM go.3SG.PST broken
‘then there comes another guy to the boy, and then the boy complains to the
guy that the bike broke down’

4.2.3 Estonian üks

Estonian üksNPs are more likely to occur with animate referents. It has
been suggested that at an initial stage of grammaticalization, ‘one’ as
an indefinite determiner is restricted to use with main characters (e.g.,
Heine & Kuteva 2006: 105). As main characters are usually animate enti-
ties, we cannot say whether animacy is actually a more influential factor

Note that the construction of the NP tälle kastelukannulla kastelijapojalle is rather
uncommon in Finnish, since Finnish does not use attributes, in particular premodi-
fiers, as productively as Estonian, for example. Yet, the speaker does not make a cor-
rection or a pause in this phrase regarding the particular example.
than the referent being a main character. However, expressing a numeral meaning is not the primary function of üks in cases like (7), where the topical referent is introduced into the story and mentioned for the first time. Such uses of üksNPs are described as the typical beginning of the narration (e.g., Pajusalu 2009).

(7) EST16, BS
ää üks noormees sõidab jalgrattaga
umm one.NOM young man.NOM ride.3SG.PRS bike.COM

‘umm a young man rides a bike’

Uses of the Estonian determiner üks are not restricted to only one character/referent in the narrative. It can be used whenever a new referent is brought into the discourse, as in Example (8). This determiner is already quite similar to an indefinite article, which is used more freely with all kinds of referents, animate or inanimate, as in Example (9).

(8) EST10, BS
ja sis tema kisa kuulis üks söber
and then 3SG.GEN yell.GEN hear.3SG.PST one.NOM friend.NOM

‘and then a friend hears his yell’

(9) EST18, KS
tuulelohe lendas tal käest ära jaa ja jää
kite.NOM fly.3SG.PST 3SG.ADE hand.ELA away and and stay.3SG.PST

kinni ühe suure kõrge puu külg
stuck one.GEN large.GEN tall.GEN tree.GEN to

‘the kite flew away from his hand and and got stuck in a large tall tree’

4.2.4 Finnish yks(i)

It has been claimed that in colloquial Finnish, an incipient grammaticalization of the numeral yks(i) ‘one’ may be in process (Schroeder 2006; see also Heine & Kuteva 2006: 127). Yet, in our data there were only 7 instances of an yks(i)NP, which exclusively referred to apple in the apple
story, as in Example (10). Therefore, while these kinds of uses are indefinite, the numeral function of yks(i) is also present in these cases (as there are actually more apples in the tree in the picture). So the presence of an indefinite article in our data is doubtful.

(10) FIN7, AS
Jaska poimii omenapuusta yhden omenan
'Jake picks one apple from the apple tree'

4.2.5. Finnish tämä

Our analysis suggests that definite tămâNPs are more common with animate entities and examples (11) and (12) demonstrate this tendency. Both examples show a tămâNP in subject position referring to a topical character. In (11), a referential distance of one utterance separates the tămâNP and the previous reference to the same entity (hän). In (12), there is another animate entity mentioned (Markus), therefore continuing with a pronoun would cause confusion. Interestingly, in both cases an adjective (nuori) is also added to the NP, which suggests that tămä again indicates an open referent which may require additional identification (Etelämäki 2006), as in Example (4).

(11) FIN9, BS
hän törmäsi kiveen. eturengas osui kiveen
3SG.NOM crash into.3SG.PST ROCK.ILL front wheel.NOM hit.3SG.PST rock.ILL
ja ja tota, sen seurauksena tämä nuori poika
and and PRTCL SE.GEN eff. tămâ.NOM young.NOM boy.NOM
lensi pyörän sarvien yli
fly.3SG.PST bike.GEN handlebar.GEN over

'he crashed into a rock. the front wheel hit the rock and and umm as a consequence the/this young boy fell over the handlebar of the bike'
HELEN HINT, TIINA NAHKOLA, RENATE PAJUSALU

(12) FIN10, BS
hän oli todennäköisesti nähny tän tilanteen
3sg.NOM be.3sg.PST probably see.PCP tämä.GEN situation.GEN
ja, ja tuli auttamaan Markusta. tään nuori
and and come.3sg.PST help.INFILL Markus.PART tämä.NOM young.NOM
mies vaihto pyörän renkaan, näppärästi paikalleen
man.NOM change.3sg.PST bike.GEN wheel.GEN handily place.ALL.Poss
‘he probably had seen this situation and and came to help Markus. the/this young man put the bike wheel handily in its place’

However, tämäNPs also are used for referring to inanimate referents as in (13); objects in the genitive or partitive cases are common in such contexts.

(13) FIN 18, AS
siellä pihallahan kasvoi omenapuu, jossa oli
there yard.ADE.PRTCL grow.3sg.PST apple tree.NOM in which be.3sg.PST
jo kolme omenaa siellä, aika lailla kypsinä ja, hän
already three apple.PART there quite much ripe.ESs and 3sg.NOM
sitten meni kastelemaan tätä omenapuuta
then go.3sg.PST water.INFILL tätä.PART apple tree.PART
‘there was an apple tree growing in the yard and it already had three apples which were pretty much ripe and he then went to water the/this apple tree’

4.2.6. A note on possessive markers

In our analysis, the group of determiners named other included an array of different determiner expressions both in Estonian (Table 4) and in Finnish (Table 5). However, within this group possessives clearly stand out as the most frequent determiner forms: 29 out of 50 in Estonian, and 25 out of 66 in Finnish. Expressing possession is structurally different in Estonian and Finnish: Estonian uses a possessive pronoun/determiner oma, while Finnish uses possessive suffixes (3sg -nsäl/-nsä, -Vn). Yet, the referential function of these forms is comparable.

94
The basic function of possessive markers is not expressing definiteness, but to denote possessive relations between entities. Nevertheless, it has been noted that third person possessive suffixes mark unique referents (Schroeder 2006: 587) and sometimes share the properties of definite articles (Gerland 2014). However, possessive markers are not directly linked to subsequent mentions, as they can be used for first introducing a referent, as well. Namely, definite possessives can refer to novel discourse entities in case of relational nouns (Barker 2000). This can be seen as one reason why conditional inference trees presented in this paper show almost equal occurrences of other NPs (including possessives) in first and subsequent mentions.

In our data, possessive NPs refer to non-subject referents, objects as in Example (14) and other syntactic roles as in Example (15), to animate as well as inanimate entities. However, more specific comparison of Estonian and Finnish possessive NPs is required in the future.

(14) EST6, KS
\text{ta sai \textbf{oma tuulelohe} tagasi}
3SG.NOM get.3SG.PST his kite.NOM back

‘he got his kite back’

(15) FIN18, BS
\text{Pietari sitten \textbf{ajoi} pyörällä\textbf{ä}\textbf{än}}
Pietari then ride.3SG.PST bike.ADE.POSS

‘then Peter was riding his bike’

5. Discussion and conclusions

In this paper we investigated how article-like definite and indefinite determiners function in two closely related (article-less) languages – Estonian and Finnish. The conditional inference tree and random forest analysis revealed that while there are important similarities between the two languages, there exist meaningful distinctions in the use of determiners.
In Estonian as well as Finnish, a very important factor that affects the choice of determiner form is number of mention, which essentially indicates the distinction between indefinite (first mention) and definite (subsequent mention) determiners.

Another factor that has an effect in both languages, but in different directions, is animacy. In Estonian, animacy only affects first-mention (indefinite) referents: the indefinite determiner üks ‘one’ is more likely to be used with animate referents. However, in Finnish, animacy is more related to subsequent (definite) mentions and shapes the choice between determiners tämä (animate entities) and se (inanimate entities) to some extent.

Two factors, syntactic role and case, are tightly connected in Estonian and Finnish, and the importance of case indicates the importance of syntactic role as well, at least to some extent. Nevertheless, Finnish seems to endorse case and Estonian favors syntactic role to explain some aspects in determiner form choice. Namely, case very clearly separates a group of determiners in Finnish, i.e., locative adverbials (e.g., sieltä ‘from there’). In Estonian, syntactic role has an effect when an NP is in subject or object position, so that these roles are more likely to occur with the definite determiner see than other syntactic roles.

The random forest analysis suggests that in both languages, age is the most important factor explaining determiner choice, although this predictor was not present in the conditional inference trees. Our data are not of sufficient size for drawing extensive conclusions concerning the influence of age, but initial observations indicate that older speakers may have fewer determiners in their speech. Previous research concerning Finnish determiners has suggested that age is indeed a relevant factor in the choice between the determiners se and tämä, so with se being more characteristic of adolescent language (Juvonen 2000; Priikki 2017). A further and more thorough analysis is needed for making more solid assessments concerning the age factor.

The present study reveals that the overall use of determiners is still relatively infrequent in Estonian and Finnish data, and NPs without
Determiners are far more common in all of the contexts described above. Using determiners is not obligatory in either Estonian or Finnish, but instead it is a pragmatic choice. On the one hand, this signals that the process of grammaticalization of definite and indefinite articles has not proceeded further than the incipient stages, as Heine & Kuteva (2006) infer. Though, the use of determiners is not random, there are specific linguistic contexts in which determiners occur and in which their main purpose is the marking of definiteness. So there is evidence that the category of definiteness is an optional grammatical category in Estonian and Finnish. This claim, concerning Finnish definite determiners, has been made earlier by Juvonen (2000: 194) who describes an optional category as less predictable and dependent on non-linguistic context, so that the use of an optional grammatical category is motivated by pragmatic or sociolinguistic factors. This is in line with Dryer’s (1983) bold claim that a word can be considered an article if it satisfies at least one criterion that defines articles. In our study, as the requirement that the word (i.e., determiner) indicates definiteness/indefiniteness, or other related discourse notions, is met, we therefore can claim the presence of optional articles in Estonian and Finnish.

Interestingly in our data, Finnish speakers use even fewer determiners than Estonian participants. Considering that earlier studies (e.g., Laury 1997; Juvonen 2000) have observed very frequent use of the determiner se in spontaneous spoken Finnish, our data indicates that language varieties also shape the use of determiners. Presumably, in a less natural experimental setting Finnish speakers tend to observe their language use more and maybe try to skip informal language elements. While Juvonen (2000) has concluded that the use of definite article-like determiners is sociolinguistically conditioned, she mostly concentrated on age as a speaker-dependent sociolinguistic factor. However, different kinds of speech situations (e.g., spoken or written language, formal or informal language etc.) also seem to have an effect on determiner use in Finnish; the exact effect of this factor remains to be analyzed in the future.
Demonstratives can also serve as placeholders (Keevallik 2010), so the question might arise how to separate article-like uses from placeholder functions. Two aspects are relevant for this study. Firstly, as Keevallik (2010) points out, there are often other constituents, such as hesitation markers, repeats, etc., between the placeholder and the actual target word, or the placeholder occurs in the middle of an NP. These kinds of uses are not accounted for as article-like determiners in our data. Secondly, an important feature of placeholders is their interactional nature: by using see as a placeholder, the speaker intends to signal to the recipient that she aims to continue speaking (Keevallik 2010). As our data contain no multi-party conversations, but only narratives, an interactional placeholder role for see is highly unlikely. It is nevertheless possible that in some occasions determiner operates as a placeholder in our data as well, since the ambiguity between determiner and placeholder functions is inevitable. However, even if the speaker has used a determiner to delay the next unit of speech, the choice of a filler (e.g., see instead of umm, repetition, etc.) must be meaningful and the determiner and placeholder co-occur simultaneously.

Acknowledgements
This research has been supported by the Estonian Research Council grant PUT701 and by the Centre of Excellence in Estonian Studies (European Union, European Regional Development Fund). We wish to thank two anonymous reviewers for their useful suggestions.

Abbreviations
In the paper appear the following abbreviations not included in Leipzig glossing rules:

- ade — adessive
- adjNP — NP with an adjective attribute
- anim — animate entity
- detNP — NP with a determiner
WITH OR WITHOUT ARTICLES?

ELA elative
ESS essive
EST Estonian
FIN Finnish
genAt genitive attribute
genNP NP with a genitive attribute
ILL illative
impl previously implicitly mentioned referent
inanim inanimate entity
INE inessive
NP noun phrase
obj object
øNP noun phrase without a determiner
PART partitive
prtcl particle
subj subject
V vocal

References


HELEN HINT, TIINA NAHKOLA, RENATE PAJUSALU


Priiki, Katri 2017. Hän, se, tää vai toi? Vuorovaikutussosiolingvistinen tutkimus henkilöviittauksista Kaakkois-Satakunnan nykypuhekelessä ['An interactive sociolinguistic study of third-person references to people in contemporary spoken Finnish in southwestern Satakunta']. University of Turku, Faculty of Humanities.


Tagliamonte, Sali A., R. Harald Baayen 2012. Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. – Language Variation and Change 24 (2), 135–178. https://doi.org/10.1017/S0954394512000129


Var teva, Annuk ka 1998. Pronominit hän ja tämä tekstissä ['Pronouns hän 'he/she' and tämä 'these' in texts']. – Virittäjä 102 (2), 202–221.


Appendix. The pictures used in the narrative elicitation task
Artikliga, artiklita? Eesti ja soome keele artiklilaadsete määratlejate võrdlus

HELEN HINT, TIINA NAHKOLA, RENATE PAJUSALU
Tartu Ülikool

Siinses uurimuses analüüsime võrdlevalt artiklilaadseid defiitseid ja indefiitseid määratlejaid eesti ja soome keeles. Eelkõige kõrveltame eesti keele defiitset määratlejat see ja soome keele defiitset määratlejaid se ja tämä ning eesti ja soome indefiitset määratlejat üks/yks(i). Samuti vaatleme eesti keele possessiivpronoomeni oma ning soome 3sg possessiivsufiksi (-nsal-nsä, -Vn) referentsiaalseid omadusi. Uurimuse põhieesmärgiks on selgitada, millised on peamised määratlejate kasutust mõjutavad keelelised faktorid eesti ja soome keeles ning kas need faktorid on keeliti sarnased või erinevad. Ühtlasi otsime vastust küsimusele, kas eesti ja soome keeles on põhjust rääkida määratlejate grammatikat慧oonist artikliteks.

Uurrat keelematerjal pärineb pildiseeria põhjal kogutud suulistest narratiividest. Uurimuses osales 20 eesti ja 20 soome keele emakeelset kõnelejat. Narratiivides esinevate viitavate määratlejaga ja määratlejata nimisõnafraaside (NP) analüüsiks rakendasime multifaktoorialeset analüüsi (tingimuslikke rekursiivseid otsustuspuid ja tingimuslikke juhumetsasid).

eelkõige subjekte ja objekte. Soome keeles seevastu on oluline kääne: omaette rühma moodustavad NP-d ablatiivi või elatiivi käändes, milles esineb sageli määreljatena lokatiivne adverbiaal.

Ehkki eesti ja soome keele määreljate kasutus ei ole juhuslik, esinevad määreljaja NP-d meie materjalis siiski oluliselt harvemini kui määreljata NP-d. Seega on määreljate grammatiseerumine artiklitest eesti ja soome keelest veel alguaamis ning definiitsuse avaldub nii eesti kui soome keelest veel selgelt pragmaatilise, mitte grammatilise kategooriana.

Võtmesõnad: määreljad; pronoomenid; grammatisatsioon; suuline narratiiv; eesti keel; soome keel

Helen Hint
Tartu Ülikooli eesti ja üldkeeleteaduse instituut
Jakobi 2, 51014 Tartu, Estonia
helen.hint@ut.ee

Tiina Nahkola
Tartu Ülikooli eesti ja üldkeeleteaduse instituut
Jakobi 2, 51014 Tartu, Estonia
tiina.nahkola@ut.ee

Renate Pajusalu
Tartu Ülikooli eesti ja üldkeeleteaduse instituut
Jakobi 2, 51014 Tartu, Estonia
renate.pajusalu@ut.ee