Pronouns as referential devices in Estonian, Finnish, and Russian

Helen Hint, Tiina Nahkola, Renate Pajusalu

Institute of Estonian and General Linguistics, University of Tartu, Estonia
School of Humanities, Tallinn University, Estonia

A R T I C L E   I N F O

Article history:
Received 25 October 2018
Received in revised form 16 September 2019
Accepted 9 October 2019

Keywords:
Personal pronouns
Demonstrative pronouns
Zero reference
Reference
Spoken narrative
Circum-Baltic languages

A B S T R A C T

This study examines the usage of different anaphoric referential pronominal expressions in three areally close languages: Estonian, Finnish, and Russian. It focuses on the linguistic factors that influence the choice of referential pronouns in spoken narrative discourse, and the relative importance of these factors in languages. We used a picture-sequence based narrative elicitation method to collect a comparable and consistent dataset for these three languages. The data is analyzed using non-parametric conditional inference tree and random forest models. The analysis shows that while the usage frequency and contexts of demonstrative pronouns show large variation in these languages, there are also devices (e.g., personal pronouns and zero reference) that act in a very similar way. There are three significant linguistic factors that influence the choice of referential devices in Estonian, Finnish, and Russian: (i) syntactic role, ii) case of the referential noun phrase, and iii) the referential distance with the previous mention. However, their exact effect, as well as the importance of other factors, varies across languages. These findings support previous research that whilst referential devices have similar grammatical labels (e.g., demonstrative pronouns and personal pronouns) in languages, their actual functions should be analyzed with respect to the specific system of the language.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

An interesting feature about language is its capability to refer to different kinds of entities by using just arbitrary sound sequences, mostly noun phrases (NPs). The traditional view of referential expressions states that different referential forms convey different cognitive status/salience/accessibility of the entities that they refer to (e.g., Givon, 1983; Ariel, 1990; Gundel et al., 1993). According to this concept, the more reduced a referential expression is, the higher its referent’s cognitive status in ongoing discourse. This means that null forms and pronouns mark the most salient and full NPs mark the less salient entities.

Abbreviations: 0, zero reference; abl, ablative case; ACC, acc, accusative case; ade, adessive case; Adv, demonstrative adverb; ALL, all, allative case; COM, com, comitative case; DAT, dat, dative case; Det, determiner; ELA, ela, elative case; Est, Estonian; FEM, feminine gender; Fin, Finnish; GEN, gen, genitive case; genAtr, genitive attribute; ILL, ill, illative case; inanim, inanimite (referent); INF, infinitive; INSTR, instr, instructive case; loc, locative case; main, main clause; MASC, masculine gender; mINF, mo-infinitive; NEUT, neuter gender; NOM, nom, nominative case; NP, noun phrase; notCl, not clear; obj, object; PART, part, partitive case; POSS, possessive suffix; PREP, preposition; PRO, PRO, pronoun; PRS, present tense; PRTCL, particle; PST, past tense; PTCP, participle; Rus, Russian; SG, singular; subj, subject; TRANS, translative case.

* Corresponding author. Institute of Estonian and General Linguistics, University of Tartu, Jakobi 2–446, 51005 Tartu, Estonia.

E-mail address: helen.hint@ut.ee (H. Hint).

0378-2166/© 2019 Elsevier B.V. All rights reserved.
Although researchers do agree that pronouns refer to salient entities in discourse, the exact nature of salience and pronoun resolution is not completely clear. This is because numerous studies have shown that the principles that the hearer uses to interpret a pronoun can be guided by various factors. Moreover, similar-looking referential devices that share the same grammatical label across languages (e.g., personal pronoun and demonstrative pronoun) may have quite different systems, discourse functions, usage contexts and referential practices in different languages.

The factors influencing pronoun use and interpretation include both sentence-level and discourse-level aspects. For example, information structure affects salience so that topics are more accessible than entities in focus or in another position (e.g., Crawley, 1986; Colonna et al., 2012). Salience is also related to the order of mention: the first-mentioned antecedent advantage (i.e., a tendency that a first mentioned referent in a sentence is more salient and thus the most probable candidate for a pronoun) has been proposed by many linguists (Gernsbacher, 1990; Järvikivi et al., 2005). Another important factor is grammatical role. Firstly, it has been noted that the pronoun as an indicator of salience will more probably refer to the subject than to any other constituent of the previous clause (e.g., Crawley et al., 1990; Grosz et al., 1995). Secondly, pronoun resolution may depend on grammatical role parallelism in that pronouns tend to refer back to the antecedents that take the same argument position in a previous utterance (e.g., Chambers and Smyth, 1998; Sauermann and Gagarina, 2017). Furthermore, some experiments have indicated the importance of the thematic role of the referent in which pronouns have a tendency to refer to the goal referent rather than to source or other roles (e.g., Stevenson et al., 1994; Arnold, 2001). It has also been suggested that salience and pronoun resolution are shaped by semantics and general world knowledge (Hobbs, 1979) or by different coherence relations (Kehler et al., 2008).

Until now, there is no clear consensus about how these different factors interact and work together. A promising attempt has been made by Kehler and Rohde (2013) to integrate coherence-based and centering-based approaches of pronoun interpretation into one probabilistic Bayesian model. However, it is not clear whether this model takes into account only English singular personal pronouns, whether it is a theory about the most reduced referring expressions only, or whether it is a more general theory of all anaphoric expressions (Kaiser, 2013). In addition, Kehler and Rohde (2013) have not included the factor of mental accessibility in their approach. This makes their model too simple to account for the real choice of referential forms in natural discourse (Ariel, 2013).

Many studies have indicated that a unified salience-based account does not always offer a sufficient theoretical base for explaining the usage-patterns of different referential devices. This is especially true in the case of personal pronouns and demonstratives in Finnish (e.g., Järvikivi et al., 2005; Kaiser and Trueswell, 2008), in Russian (e.g., Kibrik, 1996) and in German and French (e.g., Colonna et al., 2012). Therefore, they suggest that a more sophisticated form-specific multiple-constraints approach might better explain the differences in the use of referential devices since different kinds of information is used during real-time reference resolution (Kaiser and Trueswell, 2008; Kaiser, 2011). This idea has also been put forth in several earlier studies already. For example, Kibrik (1996) and Chambers and Smyth (1998) suggest that a more detailed account than a single coherence theory is needed to explain how reference resolution contributes to discourse coherence. Kibrik and his co-researchers (2016) have also acknowledged the importance of multi-factorial nature of referential choice in a study concerning machine learning and predicting referential choice, in which they list as many as 18 candidate factors of referential choice.

The present study is intended to elicit complete narrative discourse with uninterrupted and unprimed use of referential devices. In addition, our study accounts for various possible pronominal referential expressions present in the text. What we are interested in are the relations that hold between these different referential pronouns and the factors that influence the choice of a particular form. Furthermore, in this study we adopt the form-specific multiple-constraints view of the use of referential expressions. We acknowledge that referential forms with similar “tags” can have various systems and different functions and usage practices across languages. The fact that the functions of referential devices differ across languages (e.g., that some languages allow demonstratives to function as determiners while others do not) is not extraordinary per se. However, a fascinating aspect is the difference in usage practices. For example, in colloquial Finnish the demonstrative pronoun se is conventionally used as the third person personal pronoun. In Estonian, on the other hand, the demonstrative pronoun see can be used for referring to persons, but such usage contexts are more restricted than in Finnish.

By comparing areally close languages — Estonian, Finnish, and the Russian variety spoken in Estonia — we add a comparative syntactic perspective to our study of anaphoric pronominal expressions as referential devices. We examine whether the factors that influence the choice of referential expression form are similar or different in Estonian, Finnish, and Russian. More specifically, we focus on the usage patterns of categories of particular referential devices (personal and demonstrative pronouns, zero reference, demonstrative adverbs). This perspective is not specific to the study of referential devices but is relevant in linguistics more generally. For example, Haspelmath (2010) claims that cross-linguistic comparative concepts need to be distinguished from language-specific descriptive categories when conducting linguistic analysis and that system-internal investigation is necessary in order to see the exact behavior of a particular device.

First, this paper addresses the following two research questions: 1) Which linguistic factors, both grammatical (case, syntactic role etc.) or semantic/pragmatic (animacy, givenness, previous mention, referential distance etc.) guide speakers in choosing certain referential third-person anaphoric pronouns? 2) Are these linguistic factors equally important in the languages under study or is there a divergent division of labor among these factors across languages?

Second, this paper contributes to the question of how can it be attested if referential expressions with the same grammatical label (e.g., personal or demonstrative pronouns, zero reference etc.) have different discourse functions and referential practices in different languages? More specifically, we look at how to collect and analyze closely comparable data about referential noun phrases.
The following section gives an overview about the linguistic background of the languages under study and describes the systems of referential pronouns in Estonian, Finnish, and Russian. Section 3 reports on our methods of data collection and data analysis. Section 4 presents the results of our analysis together with example sentences from our data to illustrate the results. Finally, in Section 5 we summarize our work by discussing the implications of the results of this study.

2. Background: referential pronouns in Estonian, Finnish, and Russian

This study focuses on the usage of referential pronominal expressions in Estonian, Finnish, and Russian as areally and historically close Circum-Baltic languages (Dahl and Koptjevskaja-Tamm, 2001). It is important to bear in mind that this paper only deals with the Russian variety spoken by Russian people residing in Estonia and therefore it would be ill-considered to fully generalize our results to Russian spoken in Russian Federation. However, as the referential devices under study are a fundamental part of each language’s grammar, it might be expected that the tendencies of using these devices are rather similar in Russian regardless of the particular area where it is spoken.

To date, there are many studies concentrating on different third person referential devices in each language individually (e.g., Pajusalu, 2005, 2006, 2009, Reile, 2015, 2016, Hint, 2015, Vihman, 2015 for Estonian; Laury, 1997, Seppänen, 1998, Juuronen, 2000, Helasvuoto and Laitinen, 2006, Etelämäki, 2006, 2009, Priiki, 2017, for Finnish; and Kibrik, 1996, Grenoble, 1998, Nørgård-Sørensen, 1998, Prokopenya and Chernigovskaya, 2017 for Russian). In addition, there are some contrastive studies, e.g., Hint et al., 2017; Kaiser and Hiietam, 2003 for Estonian and Finnish or Pekelis (2018) for Russian contrasted to English and German. However, we are not aware of any substantial comparative studies that examine Estonian, Finnish, and Russian data in contrast in their analysis. Estonian and Finnish as Finno-Ugric languages are closely related because they belong to the same Uralic language family. Russian, on the other hand, is different because it belongs to the Indo-European languages. However, Russian shares some similarities to Estonian and Finnish regarding referential devices. Furthermore, we can expect some mutual contact-induced influence in these languages. In Estonia, there are close contacts between Estonian and the Russian variety spoken in Estonia due to extensive bilingualism and parallel use. Russian has also been a very constant contact language in Estonia in the 19th and 20th century. Estonian and Finnish as related and to some extent mutually intelligible languages are spoken in neighbouring countries and have also been in steady contact. The connection between Russian and Finnish comes down to the areal proximity, as they both belong to the Circum-Baltic region.

What is similar in all three languages regarding referring devices is that none has (fully) grammatical articles for distinguishing between definite and indefinite referents in discourse. Furthermore, due to rich inflectional case systems and subject-verb agreement, all three languages allow zero reference (i.e., an absence of overt argument in a position where it is expected) in certain contexts. However, these languages are not analyzed as full pro-drop languages; overt third person subject-verb agreement, all three languages allow zero reference (i.e., an absence of overt argument in a position where it is expected). However, Russian shares some similarities to Estonian and Finnish regarding referential devices. Furthermore, we can expect some mutual contact-induced influence in these languages. In Estonia, there are close contacts between Estonian and the Russian variety spoken in Estonia due to extensive bilingualism and parallel use. Russian has also been a very constant contact language in Estonia in the 19th and 20th century. Estonian and Finnish as related and to some extent mutually intelligible languages are spoken in neighbouring countries and have also been in steady contact. The connection between Russian and Finnish comes down to the areal proximity, as they both belong to the Circum-Baltic region.

Regarding the differences in referring strategies, Estonian and Finnish do not have grammatical gender, while Russian has masculine, feminine and neuter nouns, and pronouns. Another distinction lies in the number of demonstrative pronouns: Finnish has a stable three-demonstrative system (se ‘it’, tämä ‘this’, tuo ‘that’) and Russian has two demonstrative pronouns (eto ‘this’, to ‘that’). Standard Estonian is moving towards a one-demonstrative system (see this (and that)) from the historical three-demonstrative system (see Pajusalu, 2006). Table 1 summarizes these features in Estonian, Finnish, and Russian.

Table 1
Features of Estonian, Finnish, and Russian referring strategies.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Estonian</th>
<th>Finnish</th>
<th>Russian</th>
</tr>
</thead>
<tbody>
<tr>
<td>No (fully) grammatical article</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Grammatical gender</td>
<td></td>
<td></td>
<td>masculine, feminine, neuter</td>
</tr>
<tr>
<td>3rd person pronoun forms</td>
<td>to short ‘s/he’</td>
<td>‘han’ ‘s/he’</td>
<td>‘eto MASC ‘this’</td>
</tr>
<tr>
<td></td>
<td>remu long ‘s/he’</td>
<td></td>
<td>‘eto FEM ‘this’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>‘eto NEUT ‘this’</td>
</tr>
<tr>
<td>Number of demonstrative pronouns</td>
<td>1-term system see this, that</td>
<td>3 ‘it; s/he’</td>
<td>2 ‘eto MASC ‘this’</td>
</tr>
<tr>
<td></td>
<td>2-term system see ‘this’</td>
<td>‘tum` ‘this’</td>
<td>‘eto FEM ‘this’</td>
</tr>
<tr>
<td></td>
<td>too ‘that’</td>
<td>‘tuo ‘that’</td>
<td>‘eto NEUT ‘this’</td>
</tr>
<tr>
<td>Allow zero reference</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

1 The patterns of zero reference in Finnic languages and Russian are not completely comparable, though. Both Finnish and Estonian show a mixed pattern in the use of zero reference, allowing zero reference in the first and the second person reference and preferring an overt pronoun with the third person (e.g., Dryer, 2013). Russian, on the other hand, is rather described as a non-pro-drop language where overt pronouns are favored in all three persons (e.g., Kibrik, 1996; Dryer, 2013). As the present study relates to the usage of third person forms only, this difference is not further discussed here.

2 There is also a distal demonstrative too ‘that’ in Estonian, but its use is more restricted and characteristic to certain Estonian dialects (see Pajusalu, 2006, 2009). In our narrative data, too occurs only once.
A full description of the systems of referential devices in Estonian, Finnish, and Russian should also include demonstrative adverbs. Demonstrative adverbs frequently refer to inanimate entities which are or can be interpreted as spatial locations (Pajusalu, 2009). For example, Estonian sinna, seal, sealt, Finnish sinne, siellä, sieltä, and Russian tuda, tam, ottuda (‘to there’, ‘there’, ‘from there’, respectively) can occur in a reference chain and refer to a previously mentioned entity in text. However, demonstrative adverbs occur only in locative cases and therefore their usage contexts are more limited than those of personal pronouns, for example.

In all three languages, the boundary between demonstrative and personal pronouns can be described using syntactic features. Namely, demonstrative pronouns can be used as determiners within an NP, while personal pronouns cannot be used as determiners. Other distinctions do not work as a basis for classification: both personal and demonstrative pronouns can be used in deictic as well as anaphoric contexts, demonstrative pronouns can refer to animate and inanimate referents etc. There are differences between our languages in this regard, of course. For example, Russian demonstrative ego is rare in anaphoric use, Finnish third person pronoun hän is not used for referring to inanimate referents. The most important language-specific differences are described below.

2.1. Estonian referential pronouns

The Estonian third person singular pronoun can have either a short (ta ‘s/he’) or a long (tema ‘s/he’) form. The short form ta is a neutral choice when referring back to entities previously mentioned in an ongoing discourse (Pajusalu, 2009). The long form tema is usually pragmatically marked, for example it can express contrast to some other entity (Kaiser and Hiietam, 2003). In addition, tema more usually refers to animate entities than ta (Pajusalu, 2005).

Animate entities are usually referred to with personal pronouns in Estonian. However, in some contexts demonstrative pronouns can be used with animate\(^{3}\) entities as well (e.g., Pajusalu, 2009). Mostly, however, demonstrative pronouns refer to inanimate entities. There are two demonstrative pronoun stems in Estonian which are traditionally distinguished on a distance scale: see ‘this’ and too ‘that’ (Pajusalu, 2017: 571). However, the use of the demonstrative too in Standard Estonian is rare (Pajusalu, 2006, 2009; Reile, 2016). In addition to overt pronouns, Estonian also allows zero reference. This most commonly occurs when one referent, usually (animate) subject, is mentioned in a sequence of several utterances and there are no competing (animate) referents (e.g., Lindström, 2001; Vihman, 2015; Hint, 2015; see also Erelt, 2017: 590).

2.2. Finnish referential pronouns

The Standard Finnish third person singular pronoun is hän ‘s/he’, whereas in spoken colloquial Finnish it is very common to use the pronoun se ‘it’ for both animate and inanimate entities (e.g., Hakulinen, 1985)\(^{4}\). As Finnish has a three-term demonstrative system — the proximal demonstrative pronoun tämä (colloquial tåa) ‘this’, the distal demonstrative pronoun tuo (colloquial toii) ‘that’ and the anaphoric demonstrative pronoun se ‘it’ — there are more options for a speaker (compared to an Estonian or a Russian speaker) to distinguish between referents just by using demonstrative forms. Demonstrative pronouns can refer to both animate and inanimate entities; however, standard language has more restrictions than colloquial language in this respect (Hakulinen et al., 2008: §720). Instead of a traditional distance-based view, more recent studies of Finnish demonstratives propose an interactional interpretation for the use of different demonstrative forms (Laury, 1997; Seppänen, 1998; Etelämäki, 2006, 2009; Priiki, 2017). For example, Laury (1997) suggests that tämä refers to entities in the speaker’s sphere\(^{6}\) and tuo refers to entities outside of the speaker’s sphere, whereas se is used for referring to entities in the addressee’s sphere. It is also possible to omit the overt third person form when referring to an already-mentioned referent; this is common especially in narratives when the referent has been introduced with a lexical NP (Helasvuori and Laitinen, 2006, see also Hakulinen and Laitinen, 2008).

2.3. Russian referential pronouns

The most frequent and neutral choice of anaphoric reference to individual entities in Russian is the third person pronoun on/onə/onə ‘he/she/it’ that agrees with the gender of the previous NP referring to the same entity (Timberlake, 2004: 223; Kibrik, 2011). Similarly to Estonian and Finnish, Russian is considered a partial pro-drop language, which means that an overt phrase (noun or pronoun) is usually used for marking arguments. However, it is also possible to use the zero reference and omit overt arguments in narratives where, for example, the referent remains the same in a stretch of utterances (see Timberlake, 2004: 223–227).

\(^{3}\) In Estonian, however, animacy is usually not a grammatical but pragmatic factor. Therefore, the distinction between animate and inanimate entities in Estonian is strongly context-dependent. This is opposed to Russian, for example, where the choice of third person pronoun is strongly related to the grammatical gender system.

\(^{4}\) In this article, se is classified as a demonstrative pronoun. Depending on the sociolinguistic context, it would also be possible to classify se, when it refers to an animate entity, as a personal pronoun (e.g. Priiki, 2017).

\(^{5}\) For a definition of the concept of sphere, see e.g. Laury (1997).
There are two demonstrative pronouns used in Russian, etot ‘this’ and tot ‘that’. While Russian demonstrative pronouns are traditionally analyzed at the proximal-distal scale, this distinction holds only in contexts of juxtaposition to one another (Grenoble, 1998: 72; Timberlake, 2004: 233). The main function of demonstratives is rather presentative in that it points out entities in discourse (Grenoble, 1998, see also Timberlake, 2004). In addition, tot widely occurs in various grammatical contexts. Previous research has also shown that in Russian, basic anaphoric reference is performed by the distal demonstrative tot (e.g., Kibrik, 2011), while anaphoric contexts for the proximal demonstrative etot are very limited and rare.6

2.4. Estonian, Finnish, and Russian: a recap

Considering pronominal systems of Estonian, Finnish, and Russian, we see that each language has its own “potential advantages” in keeping discourse referents apart. First, the convenience of the Russian pronoun system lies in the possibility to distinguish different pronouns by grammatical gender. This is useful when there are competing different-gender referents in the discourse. As gender is a purely grammatical feature, interlocutors can fully rely on this distinction. Second, the Finnish pronoun system has three different demonstrative stems, which are useful when the referents are on different distance (whether spatially or in discourse). Conversely, the Finnish personal pronoun hän always refers to an animate referent, which gives it a certain referential power that Russian personal pronouns do not possess, and Estonian personal pronouns possess in a weaker form. Third, Estonian has the lowest number of different anaphoric pronouns, but, nevertheless, has two forms of personal pronoun (short and long) which could be useful when referents are distinguished by contrast.

3. Materials and methods

3.1. Data elicitation

3.1.1. Method

We used a picture-sequence based quasi-experiment to collect a well-structured and strictly comparable dataset of spoken narratives in Estonian, Finnish, and Russian.7 This method has been widely used in linguistics since the classical Pear Stories project (Chafe, 1980). We constructed three simple picture books (see Appendix 1) with six pictures in each book. The internal organization of events depicted in the picture-sequences is similar in all three sequences. There is one human character performing the action in the first and the second picture. In the third picture, another human character enters the story and becomes an active agent in the fourth and the fifth picture, while the first character has a passive role in these pictures. In the final picture, the first character is alone again.

We have included only same-gender (male) characters in all picture-sequences with the purpose of making the usage context of referential expressions more similar across languages with and without grammatical gender distinction. In addition, from each sequence we have included the three most prominent inanimate entities in our study, as we expected to have longer than one-mention reference chains formed with these entities in the narratives. Furthermore, the inanimate referents chosen can take the subject or object position and can be referred to with pronominal NPs in the narratives.

3.1.2. Participants and procedure

20 native speakers from each language, i.e., 60 volunteers in total, participated in the study. The Estonians (13 females, 7 males) were 25–47 years old (mean age 32), Finns (13 females, 7 males) were 20–80 years old (mean age 46), and Russian speakers (18 females, 2 males) were between 23 and 62 years (mean age 40).

Each participant was shown all three picture-books in a random order, and after seeing each book a participant was asked to tell a short story based on this book. To prevent extensive use of body language and text-external reference, the participants were asked to tell the story so that it would be understandable to the listeners who would only hear the story and not see the pictures. The participants were tested individually in a quiet room. The experimental sessions lasted 10–15 min and were audio-recorded.

3.2. Data coding

As each participant was asked to tell three narratives, 60 narratives in total were collected from each language. One Estonian narrative, six Finnish narratives and three Russian narratives were excluded from the final analysis due to deviations from the conventional referring strategies in narratives. These deviations included: 1) using only proper names for referring, 2) telling the story in the first and second person and using direct speech and 3) connecting the second and third narrative to the events and referents introduced in the first narrative.

---

6 We are very grateful to an anonymous reviewer for this highly valuable comment.

7 As we have published a contrastive analysis of the use of article-like determiners in Estonian and Finnish based on the same dataset, a more detailed description regarding data elicitation method in this project can be found in Hint et al. (2017).
Thus, the remaining 59 Estonian, 54 Finnish and 57 Russian narratives were coded for all referential NPs that referred to the previously selected relevant referents in the pictures (first boy, second boy and sequence-specific inanimate entities). Next, these NPs were divided into three groups:

1) Full NPs without article-like determiners (bare NPs and NPs with a genitive or adjective attribute and proper names).
2) Different pronominal forms (Pro) used as referential devices (personal pronouns, bare demonstrative pronouns, bare demonstrative adverbs and zero reference). In the case for zero reference, only third person forms occupying the subject position in the nominative case in the utterance were coded for. Although zero reference can sometimes also take other (i.e., non-subject) syntactic positions in an utterance, it is much more common with subjects than other arguments (Siewierska, 2004: 42–43). Despite there being no possibility to detect non-subject zeroes, we could determine how other factors (e.g., animacy, mention number and referential distance) could influence the choice between zero and other pronominal forms.
3) NPs with article-like definite or indefinite determiners (Det). Despite none of the three languages having grammatical articles, there are often different pronouns used in a similar referential environment to articles. These include mostly the numeral 'one', demonstrative pronouns and some other pronouns and adverbs (e.g., Estonian mingi ‘some’, Finnish siettä ‘from there’ etc.). This resulted in a total of 1304 Estonian, 1460 Finnish and 1171 Russian referential NPs.

In the statistical analysis, each referential NP was tagged for the presence/absence or features of different variables. The variables fell into two groups as follows: 1) pragmatic and semantic factors (i.e., animacy, other animate entities in the utterance, number of mention, referential distance to the previous mention, and previous NP type referring to the same entity); 2) grammatical factors (i.e., number, case of the NP, syntactic role, and clause type). These factors are summarized in Table 2.

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factors used in the study.</td>
</tr>
<tr>
<td>Factor</td>
</tr>
<tr>
<td>Animacy</td>
</tr>
<tr>
<td>OtherAnim</td>
</tr>
<tr>
<td>MentionNo</td>
</tr>
<tr>
<td>RefDist</td>
</tr>
<tr>
<td>PrevRefNP</td>
</tr>
<tr>
<td>RefNPNum</td>
</tr>
<tr>
<td>RefNPCase</td>
</tr>
<tr>
<td>SyntRole</td>
</tr>
<tr>
<td>ClauseType</td>
</tr>
</tbody>
</table>

In order to determine the number of animate entities in an utterance and to calculate the referential distance, all narratives had to be segmented into utterances. In the present paper, the notion of utterance is understood following the Centering Theory (e.g., Walker et al., 1998), since this approach has been developed specifically for explaining the underlying principles of the use of referential expressions. While our data contains spoken language, the term utterance does not indicate an intonation unit, but it is rather a fundamental segmenting unit, a tensed clause in a prototypical case. In this sense, each utterance in our analysis contains no more than one tensed predicate and all its arguments. Clause, on the other hand, is understood here as a purely syntactic concept pertaining to the description of language structure.

3.3. Data analysis

Considering our multivariate data with categorical variables, we used two non-parametric statistical methods for analyzing the data: conditional inference recursive partitioning trees (Hothorn et al., 2006) and conditional random forests (Breiman, 2001; Strobl et al., 2008; Strobl et al., 2009). The implementation on these methods for linguistic data is explained, for example, by Tagliamonte and Baayen (2012) and by Levshina (2015). Many researchers have used similar analysis for explaining different linguistic phenomena (e.g., Baayen et al., 2013; Lindström and Vihman, 2017; Levshina, 2017).

The open source statistical environment R (R Core Team, 2016) package party was used to perform the analysis.

4. Results

In this chapter, we compare the general patterns of the usage of anaphoric pronominal expressions in Estonian, Finnish, and Russian. Then, we present the results of the tree and forest analysis in each language.

---

8 We coded nominative case for the zeroes because it is possible to replace zero reference with a third person pronoun in nominative case, but not with a pronoun in non-nominative case.
9 See Hint et al., 2017 for an analysis of Estonian and Finnish article-like determiners in this data.
4.1. Referential devices in Estonian, Finnish, and Russian: an overview

For each language, we determined the overall distribution of three major types of referential devices: pronominal expressions (including pronominal adverbs and zero reference), bare NPs and NPs with a determiner (see Fig. 1).

![Fig. 1. Referential phrase types across languages (χ² = 87.699, df = 4, Cramér’s V = 0.11, p < 0.001).](image)

The most used phrase type for referring to entities in narrative discourse in Estonian, Finnish, and Russian are bare NPs, closely followed by pronouns. NPs with determiners are the least frequently used (under 20% in all languages). However, there are subtle differences between these three languages regarding the distribution of referential NPs. Firstly, Estonian has the largest proportion of detNPs (17.3% of all referential devices), closely followed by Finnish (14.4%), while in Russian, the use of determiners is only occasional (7.9%). Secondly, both Estonian and Russian have an almost equal use of bare NPs (44% in Estonian and 46.8% in Russian). Pronouns are used almost as frequently as bare NPs in Russian (45.3%), and slightly less frequently in Estonian (38.3%). But Finnish stands out in this regard as it uses pronouns much less frequently than bareNPs: the difference in usage percentages extends to more than 20 (pronouns 31.8%, bareNPs 53.8%).

The variance between the languages in their use of detNPs is meaningful, as it is stated for both Slavic and Finno-Ugric languages that there are no definite or indefinite articles (e.g., Haspelmath, 1998). However, for Estonian and Finnish, there is a line of research investigating the use of certain adnominal pronouns functioning as article-like determiners and the possibility of the grammaticalization process of articles in Estonian and Finnish (e.g., Hint et al., 2017; Juvonen, 2000; Laury, 1997; Pajusalu, 1997). Russian, at the same time, is considered as not having developed article-like uses of (in)definite determiners (e.g., Heine and Kuteva, 2006; Belaj and Matovac, 2015). One explanation comes from the fact that both Estonian and Finnish have had steady historical contact with European article-languages (German and Swedish respectively), but such strong contact with European languages has not influenced Russian historically. However, the use of pronouns as article-like determiners in Russian deserves further investigation. For example, in our data there are instances of demonstrative pronouns tot and eto functioning as definite determiners (see Example 1). Also, the Russian numeral Odin ‘one’ can be used as the indefinite determiner in some contexts (see Example 2).\(^{11}\)

\(^{11}\) For similar uses of article-like determiners in Estonian and Finnish, see Hint et al., (2017).
that curly-haired (boy) was so thankful that (he) took this/the kite.

There was a bald man with a watering can.

In total, 504 Estonian, 464 Finnish and 530 Russian pronominal expressions were coded from the data. They fell into four types as follows: personal pronouns (Prs), demonstrative pronouns (Dem), demonstrative adverbs (Adv) and zero reference (0) (see Table 3; the exact numbers regarding the presence of different pronominal forms in our data can be found in Appendix 2). Regarding Estonian, mostly personal pronouns are used (52.8%),

12 There are two possible pronoun forms in Estonian - short te and long tema (also called weak and strong forms, see e.g. Kaiser, 2010), but in partitive case, there is only one form (tede) which cannot be divided on the long-short scale. Most of the long forms in our data are in the position of a genitive attribute. While the referential functions of an NP in the position of a genitive attribute are not fundamentally different from those in some other syntactic role, we suspect that this position might influence the choice of Estonian long and short pronoun form. However, as the data are limited (there were only 8 occurrences of long forma in other syntactic environments in our data), we have not separated Estonian long and short third person pronouns in the present analysis. This is also because our aim is to compare the languages and concentrating on such subtle differences would complicate the statistical analysis.

Next, we present the results of our multivariate analysis of conditional inference trees and random forests to explain which factors influence the choice of pronoun forms in Estonian (Section 4.2), Finnish (Section 4.3), and Russian (Section 4.4).

4.2. Estonian

The conditional inference tree based on the use of Estonian pronouns in our data is presented in Fig. 2\(^1\).
The first split (Node 1) is based on the case of the pronoun. This split separates the rarely used demonstrative adverbs, marked with case tag 'other' in our data, from all other pronouns. Therefore, in every other branch of the tree, the factors describe distribution between personal and demonstrative pronouns and zero reference. In the right branch, the first distinction is made by animacy (Node 3). If there are one or more animate entities referred to in an utterance, then the tree moves on to syntactic role (Node 4). When there is only inanimate entity (or entities) referred to in an utterance, then mention number is considered important (Node 13). Here, second and further mentions after the third mention are made mostly with demonstrative pronouns, but zero forms also occur (Node 14). At the same time, third mentions and instances where the number of mention is not clear are also mostly made with demonstrative pronouns, but the zero forms closely follow, and personal pronouns are also suitable (Node 15). Therefore, for referring to already introduced inanimate entities, mostly demonstrative pronouns are used in Estonian. However, as the number of pronominal references to inanimate entities is small in our data, the distinction between Nodes 14 and 15 might not be very strong.

When the referent is animate (two boys in our picture-sequences), the importance of syntactic role becomes apparent (Node 4). Objects, genitive attributes and other non-subject syntactic roles are only referred to with a personal pronoun (Node 5), while there are more options for choosing a pronoun form in the case of given entities in the subject position. For pronominal subjects, there are two important factors: referential distance (Nodes 6 and 10) and clause type (Node 8). All pronominal forms (except demonstrative adverbs) are applicable when the referential distance with a previous mention of the same referent is 0 (the same referent is already mentioned in the current utterance) or 3 (the referent is last mentioned three utterances back) (Node 7). However, the use of personal pronouns is twice as likely here compared to demonstrative pronouns and zeroes. If the (animate) referent in the subject position is mentioned 1 or 2 utterances back, then a speaker chooses either a personal pronoun or a zero reference. Importantly, in subordinate clauses, the overt personal form is much more common than the zero (Node 9), but in main clauses (Nodes 11 and 12), overt and zero forms are in total used almost equally. However, if the referential distance with a previous utterance is 1 (the same referent is just mentioned in a preceding utterance), then the likelihood of zero form is considerably bigger than the likelihood of overt form (Node 12), but if the referent is mentioned 2 utterances back, then the preferred form is the overt personal pronoun (Node 11).

As recursive partitioning tree models can be unstable and highly variable (Strobl et al., 2009), we also ran a random forests analysis on the same data set. This is presented as a variable importance graph (see Fig. 3). The further the variable is from the vertical solid line, the more important this variable is for explaining the choice between different pronoun forms. Thus, the number of referential NP (RefNPNr), previous referential NP (PrevRefNP) and mention number (MentionNo) are considered to be unimportant variables. The three most important factors are clause type, syntactic role of the referential expression (RefNPSynt) and case (RefNPCase). Compared to the factors that are present in the Estonian tree model (see Fig. 2), mention number is deemed unimportant in the variance importance graph. This further proves the inference made above that the distinction made by mention number is not very strong (see Fig. 2, Nodes 14 and 15) in the context of this study.
An example illustrating the difference between Estonian anaphoric personal and demonstrative pronouns (see Example 3) shows that the demonstrative pronoun refers to the inanimate entity and the personal pronoun refers to the animate entity. Importantly, the personal pronoun *ta* is not the first subject pronoun in the sentence and it does not refer to the subject in the preceding utterance. Furthermore, the demonstrative *see* is used before the personal pronoun and *see* does not occur in the parallel syntactic position with the NP referring to the same entity (*kite*) in the preceding utterance.

'It's a story about a boy why flies a kite. Unfortunately, it gets stuck in a tree and he starts to cry.'

If a salient animate entity is re-mentioned in a text, Estonian speakers use either a zero reference or an overt personal pronoun. In Estonian, a zero (subject) form is syntactically mostly used in the main clause when the same entity is mentioned in a previous clause (see Example 4). More specifically, such cases very often include coordination between two clauses. While an overt personal pronoun is also acceptable under such conditions, there are more examples of overt personal pronouns when the referent is re-mentioned in a subordinate clause (see Example 5), or when the referential distance between two mentions is larger than 1, or when the pronoun occurs in a non-subject position.

![Variable Importance in Estonian](image-url)
and then the boy was happy again and rode further with the bike.

... and umm the kite-owner was happy that he got the kite back.

4.3. Finnish

The conditional inference tree describing the use of Finnish pronouns is presented below (see Fig. 4).

As in Estonian, the case of the referential NP (Node 1) separates demonstrative adverbs from other pronominal devices. In this paper, most Finnish demonstratives that are in ablative or illative case are regarded as adverbs.14 (Node 2). The occurrence of demonstrative adverbs in other branches of the tree is extremely rare and not considered hereinafter. Next, and as in the Estonian tree (see Fig. 2), the data is sorted according to whether there are one, more than one or no animate entities referred to in an utterance (Node 3). When the utterance contains reference to only inanimate entities, it is demonstrative pronouns that appear as anaphoric devices. Zero reference and demonstrative adverbs also occur, but their frequency is minor (Node 4).

---

14 Finnish has an elaborate system of demonstrative adverbs, which is partially overlapping with the locative forms of demonstrative pronouns (see e.g. Laury, 1996). As demonstrative adverbs are altogether rare in our data, we do not discuss this matter in more detail here.
When an utterance includes reference to an animate entity (or entities), then the syntactic role of the NP influences the choice of pronoun form (Node 5). As in Estonian, genitive attributes, objects and other syntactic roles are grouped into one node (Node 6) and these syntactic roles are expressed mainly with a personal pronoun or to a lesser extent with a demonstrative pronoun. Referring back to the Estonian data (see Fig. 2), this is an important difference since Estonian speaking participants did not use any demonstrative pronouns under these conditions.15

In the rightmost branch of the tree (animate subject referents), a further split is made by referential distance (Node 7). When there is a small referential distance (0 or 1), the factor of other animate entities in the same utterance is important (Node 8). Although the same pronominal forms (zero reference, personal pronoun or demonstrative pronoun) are contained in both nodes 9 and 10, the proportions of zeroes and personal pronouns are more evenly distributed when there are no other animate entities present in the utterance (Node 9). When the utterance contains references to several animate entities, the zero reference is much more frequent than the personal pronoun (Node 10). When the referential distance with the previous mention is larger than one utterance (two or more), then personal pronouns are more frequent than zero reference with demonstrative pronouns being the least frequent (Node 11).

A random forest analysis with Finnish data yielded a variable importance graph (see Fig. 5). The four factors: case (RefNPCase), other animate entities in an utterance (OtherAnim), syntactic role (RefNPSynt) and referential distance (RefDist) that are significant in the Finnish tree model (see Fig. 4) are also important in the random forest model. However, animacy is an important factor in the random forest model, but it is not depicted in the tree model. Thus, it is unclear how animacy influences the choice of referential pronoun forms in this study.

![Variable Importance in Finnish](image)

**Fig. 5.** Conditional variance importance in predicting pronoun forms for Finnish.

The Finnish analysis, similarly to the Estonian analysis, indicates that mostly demonstrative pronouns are used to indicate inanimate referents (see Example 6).

---

15 It should be noted, however, that the variation between the demonstrative pronoun se and the personal pronoun hän (see Example 8 and Example 6 respectively) when referring to an animate entity in the Finnish data has probably more to do with sociolinguistic factors than the factors taken into account in this study. Most Finnish informants who used the pronoun hän never used se in similar contexts — and vice versa.
... and my husband tasted the apple, and a smile came to his face because it was so good.

Personal pronouns and zeroes are the most frequent devices for re-mentioning the animate entities (see Fig. 4 and Examples 6 and 7). Furthermore, and as in Estonian, the Finnish zero occurs most frequently with a small referential distance from the previous mentioning of the same referent (see Example 7). Overt personal pronouns are also found within such linguistic surroundings, but to a much smaller extent.

When the referential distance is larger than one or when the (animate) entity is mentioned in a non-subject position, then the overt personal pronoun is the preferred choice. Also, demonstrative pronouns (especially se) can refer to animate entities (see Example 8). Most of our informants, however, use the personal pronoun hän (pl he), which is the default choice in Standard Finnish. Choosing hän instead of se can be interpreted as an attempt to indicate a formal register16. This attempt, in turn, is connected to the informants’ understanding of conventions related to the narrative setting (see Kalliokoski, 1991).

4.4. Russian

As indicated towards the beginning of this section (see Table 3), the Russian data differs significantly from the Estonian and Finnish data with regard to how native speakers use referential pronouns within these three languages. More specifically, Russian speakers use comparatively fewer demonstrative pronouns (1.5%) and demonstrative adverbs (0.4%) as compared to Estonian and Finnish speakers. Thus, we have excluded Russian demonstrative adverbs and demonstrative pronouns from the conditional inference tree model (see Fig. 6). This means that our Russian data analysis is conducted only on third person personal pronouns and zero reference.

Regarding the usage patterns of Russian overt personal pronouns and zero forms, only two factors turned out to be significant in the analysis: the case of the NP and referential distance (see Fig. 6). The first factor that separates the data into two is case (Node 1) so that only overt personal pronouns appear in cases other than the nominative (Node 5). When the referring act is performed in the nominative case, the choice between the overt and zero form depends on referential distance (Node 2). When the same referent is mentioned in the preceding utterance, the zero form is used in 70% of the instances (Node 4). When the referential distance is larger than one or it is unclear, the speakers use more overt personal pronouns than zeroes in our data (Node 3).

16 There are also other features of formal register in the Finnish narratives, e.g., the use of possessive suffixes and the use of expressions related to conventional story-telling (see Kalliokoski, 1991).
The Russian variable importance graph (see Fig. 7) obtained from the random forest analysis lists three significant factors for Russian: syntactic role (RefNPSynt), the case of the NP (RefNPCase) and referential distance (RefDist). Since a noun in the nominative case expresses subject in a clause in Russian (Timberlake, 2004: 276), syntactic role has clear correspondence with case in Russian. This is probably one reason why syntactic role has not been presented in the Russian tree graph (see Fig. 6).

The Russian data shows a very similar usage pattern of zero reference that was already described in Estonian (see Fig. 2) and Finnish (see Fig. 4). Namely, the zero reference is expected when the referential distance with the previous mention of the same referent is one (see Example 9). A zero form can be used with greater referential distances as well, but the proportion of the overt personal pronoun is considerably higher when referential distance is larger than one.

Fig. 6. Conditional inference tree for the distribution of pronoun forms in Russian (personal pronouns are shown in black, zero reference is shown in light grey).

The Russian variable importance graph (see Fig. 7) obtained from the random forest analysis lists three significant factors for Russian: syntactic role (RefNPSynt), the case of the NP (RefNPCase) and referential distance (RefDist). Since a noun in the nominative case expresses subject in a clause in Russian (Timberlake, 2004: 276), syntactic role has clear correspondence with case in Russian. This is probably one reason why syntactic role has not been presented in the Russian tree graph (see Fig. 6).

The Russian data shows a very similar usage pattern of zero reference that was already described in Estonian (see Fig. 2) and Finnish (see Fig. 4). Namely, the zero reference is expected when the referential distance with the previous mention of the same referent is one (see Example 9). A zero form can be used with greater referential distances as well, but the proportion of the overt personal pronoun is considerably higher when referential distance is larger than one.

(9)

<table>
<thead>
<tr>
<th>podoshel</th>
<th>prohozhii</th>
<th>pro</th>
<th>posmotrel</th>
<th>na</th>
</tr>
</thead>
<tbody>
<tr>
<td>come.to</td>
<td>passer-by</td>
<td>3SG</td>
<td>look</td>
<td>PREP</td>
</tr>
</tbody>
</table>

mač'ika | emu | stalo | ego | zhaiko |
boy | 3SG | get | 3SG | pitty |

'm a passer-by came, (he) looked at the boy, (he) got sorry for him.'

Anaphoric referential acts using non-nominative case are made by using an overt personal pronoun (see Example 10).

(10)

<table>
<thead>
<tr>
<th>kogda</th>
<th>on</th>
<th>polival</th>
<th>derevo</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>when</td>
<td>3SG</td>
<td>water</td>
<td>tree</td>
<td>PREP</td>
</tr>
</tbody>
</table>
When he was watering the tree, his friend came to him.

5. Discussion and conclusions

The main aim of this study was to explore which grammatical and semantic/pragmatic factors are involved when speakers choose various anaphoric referential expressions in spoken narratives in Estonian, Finnish, and Russian. We investigated whether these factors are, in general, similar or different across these three languages. Furthermore, we also determined whether the factors have a similar effect on the use of linguistic categories with similar labels (i.e., personal pronouns, demonstrative pronouns, demonstrative adverbs, and zero reference). The analysis of the spoken narrative data from Estonian, Finnish, and Russian revealed the following three key findings:

1) There are three important factors that help to explain the speaker’s choice of referential expression in all three languages as follows: (1) the case of the referential NP, (2) syntactic role, and (3) referential distance with the previous mention of the same referent. While the exact significance of these factors varies across languages, case is statistically the most important factor that explains certain differences in the use of referential NPs in all the three languages in our sample. There is more variation considering the order and influence of other factors.

2) The usage frequency and contexts of demonstrative pronouns is very different in our language sample. In Russian, demonstrative pronouns are not frequent referential devices in narratives, but in Finnish, demonstrative pronouns are used very often. Estonian speakers use demonstrative pronouns for referring to entities considerably less than Finnish speakers. Importantly, Estonian and Finnish demonstrative pronouns mostly refer to inanimate entities in the data. This tendency is rather straightforward in Estonian data. In Finnish demonstrative pronouns also refer to animate entities, although not that often as personal pronouns or zero forms. The few demonstrative pronouns appearing in the Russian data, however, are used for referring to animate entities.

Fig. 7. Conditional variance importance in predicting pronoun forms for Russian.
3) The three languages use similar patterns for the alternation between personal pronouns and zero reference. First, Estonian, Finnish, and Russian speakers use zero for referring to subject referents when the same referent is mentioned in the preceding utterance. Second, since overt personal pronouns are the most frequent anaphoric pronouns in both Estonian and Russian, they can be described as a default choice for anaphoric reference because there are no strict restrictions to their use. In Finnish, zero reference occurs even more often than overt personal pronouns. However, as overt demonstrative pronouns are also often used for referring to animate subjects, we can extend the ‘overt pronoun as a default’ tendency to Finnish as well. Overt third person pronouns often occur in similar contexts to zero forms (i.e., as a nominative subject in the main clause), but their usage contexts are wider. They occur more often than zeroes in subordinate clauses or when the referential distance with the previous mention of the same referent is larger than one utterance.

Regarding the overall distribution of anaphoric referential expressions in our data, the results suggest that Finnish is a demonstrative-biased language, Russian is a personal pronoun-biased language and Estonian is somewhere in between these two types. Similar results have been obtained in earlier studies, where it is also shown that the use of different demonstratives can have effect on the frequency of other grammatical constructions. For example, it has been shown that in spatial context, Finnish uses more demonstratives and less relative clauses compared to Estonian and Russian, while Russian referential practice involves demonstrative pronouns considerably less, but relative clauses are used more frequently (Pajusalu et al., 2018). Therefore, even when grammatical/referential devices in a language are named as demonstratives or personal pronouns, this does not imply that these are used in an analogous way across languages. One facet of such linguistic variety of referring expressions has already been embraced in the work of Gundel et al. (1993; see also Gundel et al., 2010) who show that while one language may encode all six cognitive statuses with a different linguistic form, there are also languages that have one single form for several statuses at once. To that end, Estonian demonstrative pronouns are mostly interpreted as referring to activated entities. Furthermore, and according to Gundel et al. (1993) original paper, Russian demonstratives esto and to also represent activated entities. Nevertheless, besides activated entities, Finnish demonstrative pronouns can also refer to entities in focus in some contexts. However, as indicated in the introduction and as our results also suggest, the cognitive status of a referential form alone does not explain its actual use in natural language discourse.

Namely, we have shown that referential forms are sensitive to different factors in different languages and that one factor can have an effect on different forms (e.g., animacy-related effects in Estonian and in Finnish). Thus, a more complex view to reference resolution, such as a form-specific multiple constraints approach (Kaiser and Trueswell, 2008) or a multifactorial concept of anaphora (Kibrik, 1996), is more plausible than a one-factor (e.g., salience) based model. Both Finnish and Estonian have been valuable model languages in this line of research. For example, research has indicated that the Finnish hän ‘s/he’ is a preferred form for referring to preceding subject referents, while the Finnish tämä ‘this’ is used with discourse-new post-verbal referents (e.g., Kaiser and Trueswell, 2008). Similarly, research concerning Estonian pronouns and demonstratives (Kaiser and Vihman, 2010) has shown that for the personal pronoun ta ‘s/he’, syntactic role is an important factor, whereas for the demonstrative see ‘this’, both syntactic role and word order factors are relevant. Our article has expanded upon these Finnish and Estonian studies by accounting for a wider range of factors (both semantic/pragmatic as well as grammatical) and illustrating the importance of a multifactorial approach (see also Kaiser, 2011; Kehler and Rohde, 2013; Kaiser and Cherqaoui, 2016). Furthermore, it is not only the individual factors, but also the combinations and relationships between these factors that influence the choice of referential expressions.

The distinct behavior of referential expressions can also be partly explained with reference to specific referential properties and usage contexts of anaphoric forms. Furthermore, we agree with previous research suggesting that referential choice is not a categorical process and in actual language use, more than one referential option are often possible (see e.g. Kibrik et al., 2016). For example, while the use of demonstrative pronouns is scarce in our data, we do not assume that demonstratives are felicitous devices of anaphoric reference in Russian. On the contrary, demonstrative pronouns are recognized as usual referential phrases in Russian (e.g., Gundel et al., 1993; Kibrik, 1996; Grenoble, 1998; Krasavina, 2011). Proximal demonstrative pronouns are even allegedly the third most frequent referential devices in Russian after full NPs and personal pronouns (Krasavina, 2011: 31–32). However, Nørgård-Sørensen (1998) has pointed out that Russian demonstrative pronouns and third person pronouns have different referential properties in that they represent different referential layers. Namely, the Russian esto ‘this’ is considered referential, but it can also occur in a text without referring to an explicit antecedent. The third person pronoun, however, must always have a well-defined antecedent, but it can also be non-referential (Nørgård-Sørensen, 1998: 176). Different layers of reference are also language-specific. In English, demonstratives and personal pronouns do not separate text world and discourse world. This has also proven to be the case in this study for both Estonian and Finnish.

Finally, besides the language’s overall system of referential devices, the choice of referential devices also depends on the referential practices, which express habitual ways of referring in different situations and general preferences of using referential devices (Hanks, 1990). In particular, the characteristics of a grammatical system of referential devices, as well as the usage frequencies of these devices, shape referential practices. For example, the grammatical gender distinction in Russian could be considered as one reason why Russian speakers rely more on the use of personal pronouns and need less demonstratives for anaphoric reference in spoken narrative context. However, Estonian and Finnish speakers do not have such
grammatical means for distinguishing between referents of different gender. Thus, in Estonian and Finnish, switching between personal and demonstrative pronouns can serve similar functions as Russian gender distinction in certain contexts.

Whilst referential practices in languages can vary considerably, it is also worth considering similar aspects of referential systems in different languages. For example, the use of zero reference is rather similar in Estonian, Finnish, and Russian. Our language sample indicates that all three languages use zero subjects often, but only when certain conditions are met. Firstly, when zero forms refer to subjects, these forms are used with a small referential distance to the preceding mention of the same referent. Furthermore, zero forms also seem to be closely connected to coordination as have been similarly observed in other languages (e.g., in Dutch, see Vliet, 2009). Due to these usage patterns, the three languages should be classified as partial null subject (or partial pro-drop) languages. This means that a zero reference is allowed in Estonian, Finnish, and Russian, but only under certain discourse conditions and in a less regular form than in “real” pro-drop languages, such as Italian (e.g., Franks, 1995; Pekelis, 2018).

A complex issue of language contacts should also be brought up with respect to the use referential devices. We cannot thoroughly look at the historical development of anaphoric referential devices in Estonian, Finnish, and Russian here, as this is not the task of this article. Moreover, the contacts are not limited to Estonian, Finnish, and Russian mutually, but there has also been a strong Indo-European influence. For example, German has influenced Estonian and Swedish has been a stable contact language for Finnish since 16th century (see e.g. Nordlund et al., 2013). However, there are some suggestions that can be made on the basis of our data. First, Estonian has probably lost its original system of demonstratives (tripartite system has turned to twopartite) and uses much more anaphoric personal pronouns than Finnish due contacts with Russian, where personal pronouns are typical anaphoric device for almost all kinds of referents. Second, NPs with determiners in our data (see also Hint et al., 2017) show that contacts are influencing languages in another way: the article-like usage of demonstratives and numeral ‘one’ is spreading from the West to the East (Heine and Kuteva, 2006: 99), in Russian this feature is not very frequent (yet).

To summarize, we have shown that the overall systems of referential expressions in Estonian, Finnish, and Russian are complex and language specific in many aspects as epitomized in the use of pronominal expressions in all these three languages. There may be substantial similarities that arise in the use of specific devices that often also share the same grammatical label in language description. However, deeper investigation of the actual language use might reveal some crucial variation in the factors that influence the use of referential devices. Thus, our results can be seen as additional evidence that unified one-factor specific explanations of reference resolution might not be adequate to capture all distinctions of different devices. Furthermore, the use of referential devices is closely related to the discourse and genre, as different devices can dominate in different genres (see e.g., Pajusalu et al., 2018). Although out of the scope of this paper, future research should be conducted into how different text types and genres may influence reference resolution and the choice of referential forms. In addition, statistical methods often neglect infrequent linguistic forms, as the analysis of Russian data demonstrates in this paper. This flaw illustrates the value of qualitative linguistic analysis. However, due to the focus of this paper and space limitations, this more thorough qualitative investigation is also left for future research.

Declaration of Competing Interest

None.

Acknowledgements

This study was supported by the Estonian Research Council, Estonia, PUT701, and by the European Regional Development Fund, European Union, Centre of Excellence in Estonian Studies. We would like to thank Roger Yallop for thorough help with language editing.
Appendix 1. Picture-sequences used for eliciting the narratives
Appendix 2. The counts of third person pronoun and demonstrative pronoun forms in the data

<table>
<thead>
<tr>
<th></th>
<th>Estonian</th>
<th>Finnish</th>
<th>Russian</th>
</tr>
</thead>
<tbody>
<tr>
<td>3SG personal pronouns</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>short ta:</td>
<td>225 (1 inanimate)</td>
<td>hän: 163 (0 inanimate)</td>
<td>on/ona/ono: 278 (45 inanimate)</td>
</tr>
<tr>
<td>long tema:</td>
<td>41 (0 inanimate)</td>
<td>total: 163</td>
<td>total: 278</td>
</tr>
<tr>
<td>total:</td>
<td>266</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demonstrative pronouns</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>see:</td>
<td>42 (40 inanimate)</td>
<td>tämä: 6 (3 inanimate)</td>
<td>tot: 8 (1 inanimate)</td>
</tr>
<tr>
<td>too:</td>
<td>1 (animate)</td>
<td>se: 109 (77 inanimate)</td>
<td></td>
</tr>
<tr>
<td>total:</td>
<td>43</td>
<td>tuo: 1</td>
<td>total: 8</td>
</tr>
<tr>
<td></td>
<td>total: 163</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References


Reile, Maria, 2016. Distance, visual salience and contrast expressed through different demonstrative systems: an experimental study in Estonian. SKY J. Linguist. 29, 63–94.


Helen Hint is writing her PhD thesis at the University of Tartu, focusing on Estonian referential devices in typological perspective. She is mostly interested in the functions and usage contexts of Estonian anaphoric referential pronouns, and in the marking of (in)definiteness and use of article-like determiners in Estonian and other languages. She is also working as a lecturer of linguistics at Tallinn University.

Tiina Nahkola received her MA in Linguistics from the University of Tampere, Finland in 2012. She is currently a visiting lecturer of Finnish language at the University of Tartu, Estonia. Her primary research interests include reference, syntax, second language acquisition and later language development. Her research is mainly focused on Finnish, her native language.

Renate Pajusalu is the Professor of General Linguistics at the University of Tartu. Her research interests include pragmatics and semantics of Estonian and adjacent languages.