A Comparison of the Referential Properties of Third Person Pronouns in Finnish and Estonian

Elsi Kaiser
University of Pennsylvania

Katrin Hiietam
University of Manchester

1. Introduction

This paper investigates the referential properties of third person anaphors in two closely related languages, Finnish and Estonian. Previous crosslinguistic research has shown that the most salient referents – i.e. referents that are most prominent at that particular point in the discourse – are referred to with the most reduced referring expressions. Moreover, factors such as (i) grammatical role, (ii) word order and (iii) the main/subordinate clause distinction have been claimed to be correlated with referent salience. In this paper, we focus on how these factors influence the referential properties of the different members of the third person anaphoric paradigms in Finnish and Estonian.

On the basis of corpus evidence, native speaker judgments and sentence completion experiments, we conclude that the referential paradigms of Finnish and Estonian cannot be mapped directly onto one another, despite some striking morphological similarities. Some referential expressions in Finnish and Estonian seem to behave in very similar ways, whereas others differ drastically. On the whole, the referential properties of the Estonian third person forms provide support for a hypothesis put forth by Kaiser (to appear) – initially formulated on the basis of data from Finnish – that we should not assume that the referential properties of all anaphoric forms can be captured in terms of a salience scale. Instead, Kaiser suggests, we need to investigate the possibility of different factors being relevant for different referential expressions. In this paper, we provide evidence suggesting that the referential properties of the Estonian third person anaphoric paradigm provide additional support for this hypothesis.

2. Third person referents

In this section we will take a detailed look at the third person anaphoric paradigm in Finnish and Estonian. Both languages have relatively free

---

1 We would like to thank the anonymous reviewer whose comments proved very helpful in preparing the final version of this paper.
word order with canonical SVO order in unmarked transitive clauses. As is typical for Finno-Ugric languages, they both have an extensive case system. Neither Finnish nor Estonian pronouns have a gender distinction, as we will see below in sections 2.1 and 2.2, and referential matrix third person subjects cannot be pro-dropped.

2.1 Finnish

The third person pronoun in standard Finnish is *hän* ‘s/he’ (ex. 1). In addition, the proximal demonstrative *tämä* ‘this’ can also be used to refer back to third person referents (ex. 2). Boldface is used to indicate coreference.²

(1) Sitten eversti piti puheen. *Hän* koetti saada ääneensä tiettyä toverillista sävyä.

‘Then the colonel gave a speech. *He* tried to get a certain friendly tone into his voice.’ (Linna:144)

(2) Lammio₁ huusi Mielosta₂, ja *tämä₂* tuli sisään lähetit kannoillaan.

‘Lammio₁ called for Mielonen₂, and *he₂* came in with the messengers on his, heels.’ (Linna:286)

The pronoun *hän* ‘s/he’ has been described as referring to the most central or ‘foregrounded’ character (Kalliokoski 1991) or to the character who is most important in a given situation or context (e.g. Vilppula 1989). According to Saarimaa (1949), *hän* tends to refer to the subject of the preceding sentence because the subject is more in the ‘foreground’ than other referents mentioned in the sentence. In contrast, the demonstrative *tämä* ‘this’, when used anaphorically, is claimed to refer to characters which are, in some sense, in the ‘background’ (see Varteva 1998). In more structural terms, Sulkala & Karjalainen (1992) claim that *tämä* ‘this’ is

---

² Abbreviations used in this paper: ACC=accusative, NOM=nominative, PART=partitive, ILLAT=illative, GEN=genitive, ADE=adessive, ALL=allative, ELA=elative, INE=inessive, COM=comitative, PRTC=participle, 3.PX=third person possessive suffix, INF=infinitive.
“used to indicate the last mentioned out of two or more possible referents” (1992:282-283).

2.2 Estonian
In Estonian there are four possible choices for third person referents. These are (i) the long and short forms of the animate pronoun tema/ta ‘s/he’, (ii) the inanimate demonstrative pronoun see ‘it, this’, which acts also as a general demonstrative in Northern Estonian dialects, and (iii) the inanimate demonstrative pronoun too ‘that’, a distal demonstrative generally in Southern Estonian dialects. The choice between see and too depends largely on the dialectal background of the speaker.

According to Pajusalu (1997), ta ‘s/he’ refers to the most prominent entity in the sentence, and the long form tema is used when the speaker contrasts a referent deictically to some other referent in the discourse. To illustrate the use of short and long forms of the third person pronouns, Pajusalu provides the example in (3). The background for this example is counting up characters in a ballet performance. In a list of characters, every new character is introduced with tema. Later, when referring back to the same person, the short form ta is used. Here, the first client has just been discussed:

(3) – ja teine klient on Maarika Aidla(.)
and second.NOM client.NOM be.3 A.NOM
	tema tanstib seltsidaami(.)
tema.NOM dance.3.SG lady companion.PART

– miks ta sinna prostituudi juurde satub (.)
why ta.NOM there prostitute.GEN at.ILL drop in.3.SG
– seda peab ise vaatama
see.PART must.3.SG REFL.NOM watch.INF
‘– and the second dancer is Maarika Aidla.’
‘– She dances the role of a lady companion’
‘– why does she drop in to the prostitute’s’
‘– that one has to watch oneself’ (from Pajusalu 1997, ex.(8)).
Now, let us consider the characteristics of the demonstratives *see* and *too* when they are used to refer back to human referents. For *see*, Erelt et al. (1993:209) state that if there are two third person referents in a clause, the pronoun *ta* is used to refer to the first referent, and *see* to the second one. It has also been noted that, in southern dialects, *too* is used for the second-mentioned referent instead of *see* (Erelt et al. 1993:209).

In sum, even though Finnish and Estonian are areally and typologically close, their anaphoric systems differ in interesting ways. In the remainder of this paper, we will investigate the referential properties of the forms summarized in Figure 1 – in particular, how they are influenced by factors including word order, grammatical role and the main/subordinate clause distinction – in order to gain a better understanding of the discourse functions of the different anaphoric expressions.

**Figure 1. Third person anaphoric paradigms of Finnish and Estonian**

<table>
<thead>
<tr>
<th></th>
<th>Finnish</th>
<th>Estonian</th>
</tr>
</thead>
<tbody>
<tr>
<td>hän</td>
<td>tämä</td>
<td>tema</td>
</tr>
<tr>
<td>see/too</td>
<td></td>
<td>see/too</td>
</tr>
</tbody>
</table>

### 3. Salience

We need to concern ourselves with the notion of salience/accessibility in order to investigate in more detail the referential properties of the anaphors in Figure 1. A general consensus is that elements with the least phonetic content refer to the most accessible or salient referents (i.e. referents that are in the current center of attention in the discourse), whereas more fully specified forms are used for less salient referents. This pattern is encoded in various accessibility hierarchies of referential forms (see e.g. Gundel, Hedberg and Zacharski 1993, Givón 1983 and Ariel 1990). Roughly speaking, according to these hierarchies, null pronouns are used for more accessible referents than overt pronouns, stressed pronouns are used for more accessible antecedents than unstressed pronouns, pronouns are used for more salient referents than demonstratives, and so on. Thus, these hierarchies would presumably predict that in Finnish, the pronoun *hän* is used for more accessible referents than the demonstrative *tämä*, and that in Estonian, the short pronoun *ta* is used for more accessible referents than the long pronoun *tema* or the demonstratives *see/too*. Moreover, at least some of the proposed hierarchies would probably also predict that the long pronoun *tema* refers to more accessible referents than the demonstratives *see/too*. However, we will in the rest of this paper that these predictions
seem to be too simple and do not match the referential patterns observed for Finnish and Estonian.

Before investigating the Finnish and Estonian third person anaphors in more detail, let us consider some of the factors that have been claimed to influence referent salience, namely (a) grammatical role, (b) word order and (c) the main/subordinate clause distinction. We chose to focus on these factors in this paper for two main reasons. First, the effects of grammatical role and word order have yielded seemingly contrasting results for different languages (as we will see below), and thus it will be interesting to see how Finnish and Estonian fit into the picture. Second, as far as we know, the role of the main/subordinate clause distinction has received less attention in the reference resolution literature and thus investigating it in parallel with the effects of word order and grammatical role can tell us more about how these different factors are related.

First, let us consider grammatical role. A significant body of research has found that subjects are more salient than objects, and that subjects are thus more likely antecedents for ‘reduced’ anaphoric forms (such as pronouns in English) than objects. For example, Crawley & Stevenson (1990) conducted a sentence completion study, where participants were given sentences followed by pronouns, as illustrated below:

(4) Paul led John along the path. He…

An analysis of the participants’ continuations revealed that the pronoun, which presumably refers to the most salient referent, tends to be interpreted as referring back to the subject significantly more often than to the object. Reading-time experiments (e.g. Gordon, Grosz and Gilliom 1993) and corpus studies (e.g. Brennan, Friedman & Pollard 1987) have also found an advantage of subjecthood.

Now, let us turn to word order. For some languages, it has been claimed that word order correlates with referent salience. For example, Strube & Hahn (1996) claim that word order influences referent salience in German, and that entities are ranked from left to right (see also Rambow 1993). However, Hoffman (1998) and Turan (1998) argue that in Turkish, salience is not influenced by word order and depends on grammatical or semantic role.

Another factor that has been associated with differences in salience is the main clause/subordinate clause distinction. Intuitively, we might expect referents in main clauses to be more salient than those in subordinate clauses. This intuition is supported by the findings of Bever & Townsend

---

4 Various other factors have also been discussed in the literature (see e.g. Arnold 1998 for a review), but for reasons of space, we will focus on only three factors in this paper.
(1979), who present experimental evidence suggesting that participants processed main clauses more deeply than subordinate clauses, and had better access to the meaning of a main clause. Related research by Miltsakaki (1999, 2002) also shows that the referential properties of entities in main and subordinate clauses pattern differently.

4. Referential properties of third person anaphora
In this section, we investigate how the three factors discussed in the preceding section, namely grammatical role, word order and the main/subordinate clause distinction, influence the referential properties of the different kinds of third person anaphors in Finnish and Estonian.

4.1 Grammatical Role
In Finnish, as mentioned above, the influence of grammatical role on the referents of the pronoun *hän ‘s/he* and the demonstrative *tämä ‘this’ was noted as early as 1949 by Saarimaa, who claimed that *hän* tends to refer to the subject of the preceding sentence and *tämä* to the object. This intuition received additional support from corpus studies by Halmari (1994) and Kaiser (2000), which show that *hän* has a distinct tendency to refer back to preceding subjects, and *tämä* prefers non-subject antecedents. However, these corpus studies were unable to disentangle the contributions of word order and grammatical role, since, in most cases, the subject linearly preceded the object (see Kaiser to appear for details). We will return to this issue below.

In light of the findings for Finnish, we would like to know whether the referential properties of the anaphors in Estonian are correlated with grammatical role. Our examination of naturally-occurring corpus data supports Erelt et al.’s (1993) and Pajusalu’s (1997) claims. As shown in ex. (5), the default referent of the short form of the pronoun (*ta*) is the preceding subject, whereas the long form (*tema*) is used when the referent is contrasted with other referents – for example, in (5), the man is contrasted with those people who will actually be using euros. These results are in line with Pajusalu’s (1997) claims that *tema* is used when contrast is involved and that the short form *ta ‘s/he* is used to refer to antecedents ‘at the center of attention.’

---

5 Corpus examples cited here are from the Estonian corpus at the Institute of Estonian (www.eki.ee/corpus), and from Õunapuu, Ervin (2000) ‘Öine fjord,’ in Novellid armastusest. Tallinn: Kirjastuskeskus OÜ.
Our corpus data show that the demonstratives (see/too) refer to non-subjects, as in examples (6) and (7). This matches observations by Erelt et al. (1993) that if there are two third person referents in a clause, see/too are used to refer to the second one.

(6) Lennart Meri₁ otsis kapten Lauri₂
L.NOM M.NOM seek.PAST.3.SG captain.ACC L.ACC
üles, sest teadis, et too₂ on töötanud.
out because know.PAST.3.SG that too.NOM be.3 work.PRTC
Kamtshatkal.
K.ADE

‘Lennart Meri₁ sought out captain Laur₂ because he knew that he₂ had been working in Kamchatka.’ (www.eki.ee/corpus)

(7) Kohtunik₁ pilgutas detektiivile₂ silma. See₂
Judge.NOM blink.PAST.3.SG detective.ALL eye.PART See.NOM
noogutas,
nod.PAST.3.SG

‘The judge₁ winked at the detective₂. He₂ nodded,…’ (Õunapuu, 2000:52)
However, as alluded earlier, it would be a mistake to now conclude that what determines the choice of hän vs. tämä in Finnish and ta vs. see/too in Estonian is the grammatical role of the antecedent. Thus far, we have not yet disentangled the contributions of word order and grammatical role, since in all the sentences that we have considered so far, the subject has preceded the object. In other words, the seemingly greater salience of subjects could be due to their linear position or their grammatical/semantic role. In the next section we separate these two factors by looking at sentences with word orders where the object occurs before the subject.

4.2 Word Order
When word order and grammatical role are disentangled, Kaiser (to appear, see also Kaiser this volume) found that, in Finnish, hän and tämä behave differently. In particular, hän appears to be primarily sensitive to grammatical role: When preceded by an SVO or and OVS sentence with two full NP arguments, hän prefers to refer to the subject regardless of the word order. In contrast, tämä is primarily sensitive to word order, and prefers postverbal referents over preverbal referents in both SVO and OVS orders. In light of these results, Kaiser concludes that we should not view the forms hän and tämä as being mapped onto a salience scale. Instead, she suggests that the two referential forms are sensitive to different factors: (i) hän cares about the syntactic function/grammatical role of potential antecedents and prefers subjects, and (ii) tämä is primarily sensitive to word order and prefers to refer to entities that are low in salience, entities that are not at the center of attention at that point in the discourse (see also Varteva 1998). In other words, the results with SVO/OVS sentences with two full NPs suggest that tämä is associated with the low-end of a salience scale, and hän with the high-end of a grammatical role scale.

For Estonian, Erelt et al. (1993) note that the pronoun ta is used to refer to the first mentioned third person referent in the preceding clause, and the demonstratives see/too to the second one. However, it is not clear whether this generalization also extends to sentences with noncanonical word order, where the object linearly precedes the subject. In fact, our small-scale preliminary survey of SVO and OVS sentences followed by ta reveals that even for OVS sentences, seven out of nine speakers (78%) interpreted ta as referring to the preceding subject. This suggests that the properties of Estonian ta might well be similar to what Kaiser (to appear)

---

6 As Kaiser (this volume) discusses, in addition to word order and grammatical role, the NP form of the antecedent is also relevant and has an impact on the referential properties of the pronoun hän.
found for the Finnish pronoun *hän*, a hypothesis that merits further research.

Now let us consider the demonstratives *see/too*. Preliminary survey data from 30 native speakers suggest that *see/too* tend to prefer the postverbal, last-mentioned referent but that they are also sensitive to syntactic role. In ex. (8), 29/20 (97%) interpreted *see/too* as referring to the postverbal object, whereas in ex. (9), 17/30 (57%) interpreted *see/too* as referring to the postverbal subject. Thus, it seems that in SVO sentences, *see/too* clearly prefer postverbal objects, but with an OVS sentence as in (13), they have only a slight preference for the postverbal subject. This is a pattern that bears some resemblance to Kaiser’s (to appear) findings for the Finnish demonstrative *tämä*, but clearly further research is needed to test the generalizability of these preliminary findings for Estonian.

(8) Mees\textsubscript{1} tervitab naist\textsubscript{2}. See\textsubscript{2}/Too\textsubscript{2} naeratab.
\begin{align*}
\text{Man.NOM accord.3.SG woman.PART. See/Too.NOM smile.3.SG} \\
\text{‘The man greets the woman. She smiles.’}
\end{align*}

(9) Naist\textsubscript{2} tervitab mees\textsubscript{1}. See\textsubscript{2}/Too\textsubscript{2} naeratab.
\begin{align*}
\text{woman.PART accord.3.SG man.NOM. See/Too.NOM smile.3.SG} \\
\text{‘It is the woman who the man greets. She smiles.’}
\end{align*}

In sum, then, these patterns suggest that just like Finnish *hän* and *tämä*, the Estonian *ta* and *see/too* differ in their sensitivity to the effects of word order and grammatical role. In Estonian, the long form *tema*, on the other hand, differs from these other options since its antecedent is whatever referent is contrastive with something else in the discourse, a property that we will return to in Section 5.

4.3 Main Clause/Subordinate Clause Distinction\textsuperscript{7}

Let us now turn to another factor that has been claimed to influence referent salience, namely the distinction between main and subordinate clauses. For Finnish, Kaiser (2000) found that in contexts where the preceding sentence contains a third-person referent in both the main clause and the subordinate clause, the demonstrative *tämä* usually refers to the referent in the embedded clause (ex. 10), and the pronoun *hän* to the referent in the main clause (ex.11).

\textsuperscript{7} There are various kinds of subordinate clauses in a language, e.g. report complements, relative clauses, temporal adjuncts etc. As in Kaiser (2000), at this stage we have not yet distinguished between different subtypes of subordinate clauses here and treat them as one.
Deep in thought, the sergeant looked towards the vestibule, where the scribe was combing his hair. He was making faces in front of the mirror. (Linna:23)

Lewinsky nodded, chewed her lips, tossed her hair around, and smiled, when the interviewer got mixed up in his words. Most of the time she seemed a little bit bored. (Helsingin Sanomat, Verkkoliite, 7/2/99)

Additionally, Kaiser (2000) found that the demonstrative was used in certain contexts in which a pronoun would in fact have been unambiguous. These tended to be contexts where the antecedent of the demonstrative was in an embedded position (e.g. the object of a preposition in sentence such as ‘The apple landed a few feet from Bill’) and where another referent in that same sentence was more prominent at that point in the discourse. In light of these findings, Kaiser (2000) suggests that the demonstrative tämä can be used to when its antecedent is significantly less salient than some other entity in the surrounding context.

Let us now consider how the main clause-subordinate clause distinction is reflected in the choice of anaphora in Estonian. A survey conducted with 30 native speakers shows that ta tends to be interpreted as usually referring to the subject of the matrix clause, as in example (12) (22/30, 73% matrix-subject interpretations). In contrast, see/too take the subject of the subordinate clause as their antecedent, as in example (13) (27/30, 90% subordinate-subject interpretations).
A COMPARISON OF THE REFERENTIAL PROPERTIES OF THIRD PERSON PRONOUNS

(12) Liisa₁ märkas, et Mati₂ seisis uksel.
    L.NOM notice.PAST.3.SG. that M.NOM stand.PAST.3.SG door.ADE.
    Ta₁ naeratas.
    3.SG.NOM smile.PAST.3.SG
    ‘Liisa₁ noticed, that Mati₂ was standing at the door. She₁ smiled.’

(13) Liisa₁ märkas, et Mati₂ seisis uksel.
    L.NOM notice.PAST.3.SG. that M.NOM stand.PAST.3.SG door.ADE
    See/Too₂ naeratas.
    3.SG.NOM smile.PAST.3.SG
    ‘Liisa₁ noticed, that Mati₂ was standing at the door. He₂ smiled.’

These data suggest that ta prefers to refer to the subject of the matrix clause, and the demonstrativess to the subject of the embedded clause. This matches the pattern observed for hän and tämä in Finnish. Let us now consider what happens if the short form ta in (12) is replaced with the long form of the pronoun, tema. Recall that the long form is used when the referent contrasts with something else. When presented with the sentence out of context, the questionnaire participants’ responses were split between the matrix subject (18/30, 60%) and the embedded subject (12/30, 40%). This is not surprising, since either one of these subjects could be construed as contrastive, and the bias towards the main subject could be explained as a result of matrix subjects being more salient than embedded subjects.

In sum, we can say that in terms of their tendencies to refer to entities in main or subordinate clauses, the short pronoun ta seems to pattern like the Finnish pronoun hän, and the demonstratives see/too resemble the Finnish demonstrative tämä. Interestingly, the long pronoun tema in Estonian is sensitive to contrast and is not used to refer to less salient referents – even though this is what one might have predicted in light of the claims made by accessibility hierarchies.

5. Discussion and conclusions

In this paper we investigated the referential properties of third person referential expressions in Finnish and Estonian (Fig 1, repeated below); in particular, we were concerned with the effects of word order, grammatical role and the main/subordinate clause distinction on the use of these referential forms.
Figure 1. Third person anaphoric paradigms of Finnish and Estonian

<table>
<thead>
<tr>
<th>FINNISH</th>
<th>ESTONIAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>hän</td>
<td>ta</td>
</tr>
<tr>
<td>tämä</td>
<td>tema</td>
</tr>
</tbody>
</table>

We used corpus evidence and native speaker survey data for Estonian to investigate the referential properties of the Estonian forms and to compare them to the patterns observed for the Finnish pronoun hän and demonstrative tämä. Our preliminary results suggest that the Estonian short pronoun ta may pattern like the Finnish pronoun hän, and the Estonian demonstratives see/too seem to act like the Finnish demonstrative tämä. Interestingly, the Estonian long form tema, which is historically related to the Finnish demonstrative tämä (Kulonen et al. 2000:355), has a very different function in Estonian than tämä does in Finnish. In Estonian, tema is used to refer to entities that are contrasted with something else (e.g. Pajusalu 1997), whereas in Finnish tämä is used for entities low in salience.

This striking difference in the discourse functions of the two closely related forms tema and tämä – as well as the observation that ta and see/too seem to be sensitive to different kinds of factors – deserves further research and supports Kaiser’s (to appear) claim that we should not assume that the different members of the anaphoric paradigm of a given language can be mapped along a salience/accessibility scale. As mentioned earlier, accessibility hierarchies that have been proposed in the literature would presumably predict that in Finnish, hän is used for more accessible referents than tämä, and that in Estonian, ta is used for more accessible referents than tema or see/too. Moreover, at least some of the hierarchies probably also predict that tema refers to more accessible referents than see/too. However, as we have tried to illustrate in this paper, in our investigation of the Finnish and Estonian anaphoric paradigms, we have not found these patterns. In sum, our preliminary findings support Kaiser’s claim that we need to explore the idea of different factors, such as grammatical role, discourse-status and contrast, being relevant for the different members of an anaphoric paradigm (see also Kaiser (in preparation)). We plan to investigate this hypothesis in more depth in future work.
A COMPARISON OF THE REFERRENTIAL PROPERTIES OF THIRD PERSON PRONOUNS

References:
Gordon, Peter C., Grosz, Barbara J., & Gilliom, Laura A. 1993. ‘Pronouns, Names, and the Centering of Attention in Discourse,’ Cognitive Science 17, 311-347


