

Effects of Contrast on Referential Form: Investigating the Distinction Between Strong and Weak Pronouns

Elsi Kaiser

*Department of Linguistics
University of Southern California*

To further our understanding of the nature of the form–function mapping in anaphoric paradigms, this study investigated the referential properties of strong pronouns (long pronouns) in Estonian. Cross-linguistically, 2 main accounts of the long–short distinction have been proposed: the salience account (long pronouns refer to less salient antecedents) and the contrast account (long pronouns refer to entities that are being mentioned contrastively). To test these claims, this study compared parallel corpora of Estonian and Finnish to see how Estonian long pronouns are realized in Finnish and what the grammatical role of the antecedent is. Building on Pajusalu (1997), this study also analyzed the referential properties of long pronouns from the perspective of alternative semantics (Rooth, 1992) and Jackendoff’s (1972) and Büring’s (2003) research on contrast. The corpus patterns support the contrast account, indicating that the long–short distinction cannot be straightforwardly reduced to referent salience. As a whole, these results fit with the *form-specific multiple-constraints* approach to reference resolution (Kaiser & Trueswell, 2008).

Successful linguistic communication requires the ability to refer to already-mentioned entities and events, and to correctly interpret such references when they are produced by others. This information-management process often involves semantically impoverished forms like pronouns (e.g., *she* or *it*) and

Correspondence concerning this article should be addressed to Elsi Kaiser, Department of Linguistics, University of Southern California, 3601 Watt Way, GFS 301, Los Angeles, CA 90089–1693. E-mail: emkaiser@usc.edu

demonstratives (e.g., *this* or *that*). It has been widely assumed that our ability to interpret these “shorthand” forms depends on a particular form–function mapping. In particular, researchers have argued that referential forms can be ranked along a salience scale, with the most-reduced referential forms (e.g., pronouns in English) being used to refer to entities that are most prominent or most important at that point in the discourse, whereas other forms (e.g., demonstratives or full noun phrases [NPs]) are used for increasingly less prominent, less salient entities.

However, recent cross-linguistic work suggests that not all referential forms map onto a unified salience scale, and that the salience–prominence of the intended referent may not be sufficient to explain the use and interpretation of various anaphoric forms (see Kaiser, 2003; Kaiser & Trueswell, 2008; see also Brown-Schmidt, Byron, & Tanenhaus, 2005; Kaiser, Runner, Sussman, & Tanenhaus, 2009). In fact, findings from a range of languages support the *form-specific multiple-constraints* approach (Kaiser & Trueswell, 2008), according to which different referential forms exhibit varying levels of sensitivity to different aspects of the antecedent—in other words, each form has its own set of weighted constraints that guide its interpretation. As a result, according to this view, languages can contain forms whose referential properties cannot be straightforwardly reduced to a single salience dimension.

The research reported in this article has two interrelated goals. First, by looking at the pronominal and demonstrative system of Estonian and comparing it to Finnish, I aim to further our understanding of the form–function mapping in anaphoric paradigms; and, thereby, to address open questions that have implications for the form-specific multiple-constraints approach. Second, in doing this, I also aim to shed light on a typologically well-documented, but theoretically rather understudied, linguistic phenomenon—namely, the distinction between two types of overt pronouns: so-called weak or short pronouns and strong or emphatic (long) pronouns. In particular, the discourse functions of emphatic pronominal forms are not yet well-understood, and existing research on this topic has resulted in divergent accounts.

In Estonian, third-person human referents can be referred to with the long pronoun *tema*, as well as with the short pronoun *ta* (see Example 1). Both of these forms are gender neutral:

1. {*Ta/Tema*} *hakkas naerma*. ‘She/he started to laugh.’

Cross-linguistically, two main accounts of the long–short distinction have been proposed: the *salience account* and the *contrast account*. Although other possible explanations of this distinction are also possible, these two accounts have received the most attention in existing research. The salience account posits that use of long forms is sensitive to referent salience, and that long/emphatic

forms are used for lower-salience referents than short forms (e.g., Cardinaletti & Starke, 1999). According to the contrast account, long forms are used for referents that contrast with other discourse entities (e.g., Pajusalu, 1995, 1997). Both accounts agree that short forms are used for highly salient/topical antecedents.

The corpus studies presented here investigate these accounts by looking at the distribution and referential properties of the long pronoun *tema* in Estonian and comparing parallel corpora of Finnish and Estonian, two closely related Finno-Ugric languages. The pairing of Finnish and Estonian is especially useful for shedding light on the referential properties of long pronouns due to these languages' unique combination of similarities and differences. Finnish and Estonian both use personal pronouns ("She/he smiled") and demonstratives ("This smiled") to refer to human antecedents, and it is widely agreed that demonstratives are used for relatively lower-salience antecedents in these languages. However, in Finnish, there is only one form of the gender-neutral third-person pronoun (*hän* 'she/he'), whereas Estonian has a long/emphatic pronominal (*tema*) and a short, default pronominal (*ta*). I use this asymmetry to investigate the referential properties of the Estonian long form *tema*.

On the basis of corpus data, I suggest that the salience account is not sufficient and that use of the long form *tema* is sensitive to the presence of contrast, as suggested by Pajusalu (1995, 1997). Furthermore, by investigating the referential properties of *tema* from the perspective of alternative semantics (e.g., Rooth, 1992), as well as Jackendoff's (1972) and Büring's (2003) research on contrast, I build on Pajusalu's (1995, 1997) observation by providing a more explicit discussion of the discourse contexts in which *tema* occurs. The significance of this research goes beyond Estonian, as the results presented here have implications for our understanding of whether the use and interpretation of referential forms is best conceptualized as driven by a single, unified dimension such as salience, or whether it is a more multidimensional process in which referential forms differ as to which factors they are most sensitive to.

EXISTING RESEARCH

The nature of the mapping between referential forms and their intended antecedents has received considerable attention. Given the limited semantic information carried by forms such as *she* or *this*, how do comprehenders understand what entity the speaker intends to refer to? A large body of research supports the idea that the form of referring expressions is connected to the salience of their referents. The term *salience* refers to how prominent or how foregrounded different entities are in the language users' mental model of the discourse at a particular point in time. The claim is that the most-reduced referring expressions refer to highly salient, highly prominent referents, and that fuller expressions are

used for referents that are less salient and less central (e.g., Ariel, 1990; Givón, 1983; for a different, cognitive-status-based approach that uses an implicational scale, see Gundel, Hedberg, & Zacharski, 1993). Part of a generalized salience hierarchy is shown in Example 2, where the forms that are further to the left are used for more salient referents. Researchers who work within this approach commonly agree that demonstratives are used for less salient referents than pronouns (e.g., Ariel, 2001, p. 29).

2. *null > pronoun > demonstrative > full NP ...*

The distinction between short and long pronouns has received considerably less attention than the null versus overt pronoun distinction (e.g., on null and overt pronouns in Spanish and Italian, see Carminati, 2002; Luján, 1985) or the pronoun versus demonstrative distinction (e.g., Bosch, Rozario, & Zhao 2003; Kaiser, 2003; Kaiser & Trueswell, 2008; Kibrik, 1996). However, the underlying assumptions of many salience-hierarchy approaches lead to the prediction that long forms are used for less salient antecedents than short forms. Ariel (2001) argued that “attenuation” in terms of “phonological size” is one of the key criteria that determines what degree of accessibility the antecedents of a particular referring expression have (p. 32). More attenuated, phonologically reduced referring expressions have more salient antecedents than less-reduced referring expressions. Thus, the prediction is that the long form *tema*, by virtue of being long, has less salient antecedents than the short form *ta*. The view that attenuation/reduction is related to salience also fits well with other work showing that highly predictable material can be reduced in production (e.g., Gahl & Garnsey, 2004; Jurafsky, Bell, Gregory, & Raymond, 2001), especially when combined with Arnold’s (1998) claims that salience is connected to likelihood of upcoming mention.

In related work, Cardinaletti and Starke (1999) argued that strong pronouns and weak pronouns (which they called deficient) differ in terms of their morphosyntactic structure and the prominence/salience of their antecedents. According to their view, strong pronouns “are able to refer to a non-prominent discourse referent” (p. 154) in contrast to weak pronouns, which “must have an antecedent prominent in the discourse” (p. 154). Although Cardinaletti and Starke did not discuss Estonian in particular, *tema* and *ta* fit their definitions of strong and weak pronouns, respectively. Thus, their approach predicts that long forms like *tema* have less salient antecedents than short forms like *ta*.

Assuming that (a) pronouns as a group refer to more salient antecedents than demonstratives (following Ariel, 2001, and others) and that (b) long and short forms count as referentially distinct forms and occupy different positions in the hierarchy, the following hierarchy is predicted. (Forms to the left refer to more salient antecedents than those to the right):

3. *reduced/short pronoun* > *full/long pronoun* > *demonstrative* > *full NP* ...

Before turning to the other account, according to which use of long forms is sensitive to the presence of contrast, let us consider one of the key issues related to the salience account—namely, the question of what influences a referent's salience. Existing research suggests that a range of factors play a role (for an overview, see Arnold, 1998), including grammatical role (e.g., Brennan, Friedman, & Pollard, 1987; Chafe, 1976; Crawley & Stevenson, 1990), verb semantics (e.g., Stevenson, Knott, Oberlander, & McDonald, 2000), linear order (e.g., Gernsbacher & Hargreaves, 1988), and structural parallelism (e.g., Chambers & Smyth, 1998). An important, related issue concerns the effects of cross-clausal semantic relations on pronoun interpretation (e.g., Kehler, 2002; Kehler, Kertz, Rohde, & Elman, 2008). Perhaps the most often-observed finding—and the one that is most central to this research—concerns grammatical role: A sizeable body of research shows that entities in subject position are more likely to be the antecedents of subsequent (subject) pronouns than non-subject entities.

However, despite the large body of work connecting referential forms and salience, not everyone agrees that the salience account is sufficient to explain the use of long versus short pronominal forms. Even if one assumes that referential forms are ranked along a salience scale, there is still the question of whether (a) *ta* and *tema* are regarded as two distinct elements that can be ranked differently on the hierarchy or whether (b) they are treated as two versions of the same pronominal form and, therefore, situated on the same level in the hierarchy.

If *ta* and *tema* occupy different positions on the hierarchy, *ta* is predicted to refer to more salient referents than *tema*. However, if *ta* and *tema* are situated at the same level on the salience hierarchy, the choice of *ta* versus *tema* would presumably be determined by some other factor, perhaps something like contrast. Existing research on Estonian lends support to this idea. On the basis of corpus data, Pajusalu (1995, 1997) suggested that the use of *tema* is triggered by comparison or contrast. Similar observations have been made for Dutch by Haeseryn, Romijn, Geerts, de Rooij, & van den Toorn (1997, p. 252; see also Kaiser, 2003; Kaiser & Trueswell, 2004, on the interpretation of Dutch anaphors). Further evidence for the role of contrast comes from the null versus overt pronoun distinction in Spanish (e.g., Luján, 1985; but, for a topicality-based account, see Alonso-Ovalle, Clifton, Fernández-Solera, & Frazier, 2002).

In sum, on the basis of existing research, we can formulate two main views regarding the referential properties of long versus short pronouns: (a) the *contrast approach*, which claims that the long form is used when the antecedent contrasts with another referent in the domain of discourse; and (b) the *salience approach*, which claims that the long form is used to refer to less salient antecedents than the default short form. It is worth noting that these accounts are not

necessarily mutually exclusive, as it could turn out to be the case that both salience differences and the presence or absence of contrast play a role in the long–short distinction. In light of this, the corpus analyses presented in this article investigate both dimensions to see if one or both are relevant.

FORM-SPECIFIC MULTIPLE-CONSTRAINTS APPROACH

The question of whether the choice of *ta* versus *tema* can be entirely captured on the basis of a salience scale—or whether another notion, such as contrast, is necessary—has broader implications that reach beyond these particular forms in Estonian. In particular, challenging the widespread view that different referential forms can be mapped onto a unified salience scale, Kaiser (2003) and Kaiser and Trueswell (2008) proposed a new account of reference resolution, according to which anaphoric forms can differ in how sensitive they are to various antecedent properties. On the basis of psycholinguistic experiments on Finnish pronouns and demonstratives, Kaiser and Trueswell (2008; see also Kaiser, 2003) proposed the *form-specific multiple-constraints approach*, according to which anaphoric forms can differ in how sensitive they are to different properties of the antecedent. According to this account, each anaphoric form has its own set of weighted constraints that shape its referential properties. For example, one form may be very sensitive to the grammatical role of potential antecedents but less sensitive to another factor, such as the presence of contrast; whereas another form may be more sensitive to a factor, such as contrast, and relatively insensitive to grammatical role. Using sentence completion and eye-tracking methods to investigate reference resolution following sentences with subject–verb–object (SVO) and object–verb–subject word order, Kaiser (2003) found that, whereas Finnish pronouns were primarily sensitive to the antecedent’s grammatical role and preferred subjects (see also Järvikivi, van Gompel, Hyönä, & Bertram, 2005), demonstratives were sensitive to linear order, as well as grammatical role, preferring post-verbal antecedents, especially objects (see also Kaiser & Trueswell, 2008). This asymmetrical sensitivity is problematic for an approach that aims to rank all referential forms on a unified salience scale.

Further evidence for the form-specific approach comes from English pronouns and demonstratives, investigated by Brown-Schmidt et al. (2005). Using sentences such as, “Put the cup on the saucer. Now put it/that . . .,” Brown-Schmidt et al. found that *it* was more sensitive to the antecedent’s grammatical role than *that*, which preferred the composite (cup + saucer). In recent work on English pronouns and reflexives, Kaiser et al. (2009) showed that the form-specific approach extends to within-sentence reference resolution. Work

on Dutch weak and strong pronouns by Kaiser and Trueswell (2004) suggested that a unidimensional salience account may also be insufficient for Dutch.

Investigating whether the form-specific approach extends to Estonian long and short pronouns contributes to our understanding of reference resolution in several ways. First, so far, the focus of form-specific research has mainly been on referential forms that are clearly separate lexical items (personal pronouns–demonstrative pronouns and pronouns–reflexives). By looking at long and short pronouns in Estonian, we can test whether two related anaphoric forms that are morphologically distinct but conceptualized by language users as being two variants of the same word can also differ in how sensitive they are to different antecedent properties.

In addition, this work aims to further our understanding of what kind of contrast may be relevant for the use of long pronouns. Although the intuitive notion of contrast was mentioned by Pajusalu (1997) in her earlier work on long and short Estonian pronouns, she did not offer an explicit definition or formalization of this term. One of the key aims of this work is to apply existing (non-pronoun related) research on contrast—in particular, theoretical work on contrastive focus (in the sense of Rooth, 1992) and contrastive topics (in the sense of Büring, 2003)—to the distribution of Estonian long pronouns. This will allow us to take steps toward a clearer understanding of whether and how contrast and the presence of alternative referents influence use of strong pronouns. Existing research within the form-specific multiple-constraints framework has primarily focused on other factors (e.g., sensitivity to grammatical role and sensitivity to linear position), and the issue of whether different anaphoric forms may differ in how sensitive they are to contrast is not yet well-understood.

Finally, by comparing Estonian to Finnish, we can further our general understanding of morphologically complex anaphoric paradigms and the extent to which the form–function mapping differs among languages that are typologically distinct from English. In sum, the question of how well the salience account and the contrast account capture the referential properties of long and short pronouns will contribute not only to our understanding of Estonian, but also has broader implications for our view of reference resolution and the extent to which it is guided by form-specific information.

BACKGROUND ON ESTONIAN AND FINNISH

Estonian and Finnish both belong to the Finnic branch of the Uralic language family, and have rich case-marking systems and flexible word order. In both languages, the canonical word order is SVO, but all six configurations of subject, verb, and object are felicitous under the right discourse conditions (on Estonian, see Tael, 1988; on Finnish, see Vilkkuna, 1989).

Estonian Anaphoric Paradigm

In Estonian, there are three anaphoric choices for human third-person referents: (a) the long form of the gender-neutral pronoun *tema* ‘she/he’; (b) the short, default pronominal form *ta* ‘she/he’; and (c) the general demonstrative *see* ‘this’. According to Pajusalu (1995, 1997) and many others, the short pronominal form *ta* ‘she/he’ is the default third-person pronoun in Estonian, and functions as a means of referring to the entity in the focus of attention (Pajusalu, 1997, p. 107; see Example 4):

4. [Context: debate about whether stills from a home movie can be published]

Ta ütles, et ta on autor ja tal on autoriõigused ja et ta teeb, mis ta tahab.
 ‘He said that he is the author and he has the author rights, and that he will do what he wants’ (see Eesti Ekspress Online at <http://www.ekspress.ee>).

Estonian also has other ways of referring to human antecedents, including the demonstrative *see* ‘this’. As in English, this demonstrative can be used in a range of ways, including deictically as a proximal demonstrative (e.g., “Hey, look at this!”), as a discourse deictic (e.g., “Peter pushed John. This was a mean thing to do.”), and as a pronominal modifier (e.g., “this man”). In addition, unlike English, *see* ‘this’ can be used to refer to human antecedents. It is this anaphoric use of *see* that is of interest to us in this article. (Our discussion of demonstratives focuses mainly on *see*, which is used as a general demonstrative in northern Estonian [Pajusalu, 1996], widely regarded as standard Estonian.)

According to Erelt et al. (1993, p. 209) and Tauli (1983, p. 323), if there are two third-person human referents in a clause, *ta* is used to refer to the first-mentioned referent and *see* to the second-mentioned referent (see Examples 5a & 5b; subscripts indicate co-reference). In certain dialects of Estonian, especially in the south, the demonstrative *too* is used anaphorically in addition to, or instead of, *see* (Erelt et al., 1993, p. 209). According to Erelt et al., *too* and *see* are used in the same way (see Example 6); and, in this article, I group them together (see also Pajusalu, 1997, p. 114):

- 5a. *Ta₁ vaatas tüdrukule₂ enda ees. See₂ oli kodune ja lihtne tütarlaps.* ‘He₁ looked at the girl₂ in front of him. She₂ was a homely and simple girl’ (Tauli, 1983, p. 323).
- 5b. *Tüdruk₁ vilksas poisile₂ poole; ta₁/see₂ oli kahvatu.* ‘The girl₁ glanced toward the boy₂; she₁/he₂ was pale’ (Erelt et al., 1993, p. 209).
6. *Lapsed₁ polnud oma isa₂ näinud ja ei teadnud, kuhu too₂/see₂ oli kadunud.* ‘The children₁ had not seen their father₂ and did not know where he₂ had disappeared’ (Erelt et al., 1993, p. 209).

Before considering some small-scale corpus findings regarding *see*, it is worth noting that Example 6 suggests that use of *see/too* is not triggered by a need to disambiguate between two competing referents. The subject in Example 6 is plural (children) and, thus, use of a third-person singular pronoun *ta* would have been unambiguous (for related data from Finnish, see Kaiser, 2005).

The claim that the demonstrative *see* prefers objects is corroborated by preliminary small-scale corpus studies: We found that the short form *ta* has a strong subject bias (>70% refer to matrix subject), and the demonstrative *see* has a strong preference for antecedents that are realized in direct object position or in more oblique positions (e.g., indirect objects, locatives, and other adjunct-type positions); over 90% of the occurrences of *see* refer to objects or obliques (Kaiser, 2003; see also Kaiser & Hiietam, 2004). If we assume that subjects are more salient/prominent than entities realized in other positions, then Estonian fits with the salience hierarchy-based prediction that pronouns refer to more salient antecedents than demonstratives.

As a slight aside, it is worth noting that the situation is actually more complex because word order also matters: Kaiser and Vihman (in press) and Kaiser and Hiietam (2004) found that to fully capture the referential properties of *ta* and *see*, we need to take into account both the grammatical role and the linear position of the antecedent. However, our primary focus here is on sentences where subjects precede objects (i.e., grammatical role and linear order coincide); thus, this research does not distinguish word order from grammatical role.

In addition to *ta* and *see*, the long pronominal form *tema* is also used to refer to human antecedents. On the basis of a corpus study, Pajusalu (1997; see also Pajusalu, 1995) suggested that *tema* is used to refer to entities that are being opposed to or compared with another referent in the domain of discourse, as in Example 7 (from Pajusalu, 1997, p. 109). (I focus here on nominative occurrences of *tema* in subject position. As Pajusalu [1997] pointed out, the *ta/tema* alternation is more restricted for oblique cases, but both forms can occur in subject position.) Here, the different dancers in a ballet are being announced. According to Pajusalu (1997), the speaker uses *tema* when specifying the second dancer's role because it is being compared to the first dancer's role:

7. *-ja teine klient on Maarika Aidla. Tema tantsib seltsidaami.* ‘-and the second dancer is Maarika Aidla. She dances the role of a lady-in-waiting.’

However, because it was not the central focus of her study, Pajusalu (1997) included some corpus examples of *tema*, but did not provide an explicit definition of *contrast*. Given that different researchers have defined the concept of contrast in different ways (for an overview, see Molnár, 2006), one of the key aims of this article is to analyze the possibility of contrastive usage of Estonian *tema* in more detail.

It is important to note that the *taltema* distinction does not clearly map onto an accented/unaccented pronoun distinction: Whereas *ta* cannot be accented, *tema* can be used in an accented or unaccented position (Pajusalu, 1995, citing Palmeos, 1981; Vseviiov, 1983). Furthermore, even if the *taltema* distinction were to correlate with the presence/absence of accent/stress, this would not necessarily be unambiguous evidence in favor of the contrast account: Existing work on stressed pronouns in English is characterized by conflicting views. Some researchers (e.g., de Hoop, 2003) argued for a contrast-based account, and others (e.g., Ariel, 1990; Kameyama, 1999) adopted a salience-oriented approach. Thus, even if we were to regard the Estonian *taltema* distinction as directly mapping to an accented/unaccented distinction, existing research does not provide an unequivocal answer regarding the roles of contrast and salience in guiding the choice of the long versus short form.

Finnish Anaphoric Paradigm

Previous work on the Finnish personal pronoun *hän* ‘she/he’ describes it as referring to the most central, foregrounded character (e.g., Kalliokoski, 1991) or to the most important character in a given situation or context (e.g., Vilppula, 1989). Saarimaa (1949) suggested that *hän* tends to refer to the preceding subject, as shown in Example 8 from V. Linna’s (1954/1999) novel, *Tuntematon Sotilas* ‘Unknown Soldier’. Unlike the long/short distinction exhibited by Estonian *temalta*, the Finnish personal pronoun *hän* does not have a longer or shorter counterpart. Like Estonian, standard Finnish also uses the demonstrative *tämä* ‘this’ anaphorically for human antecedents (see Example 9; on referential patterns in Finnish dialects, see Seppänen, 1998). As with Estonian *see*, the core function of *tämä* is demonstrative; it can be used deictically and as a pronominal modifier:

8. *Sitten eversti₁ piti puheen. Hän₁ koetti saada ääneensä tiettyä toverillista sävyä.* ‘Then the colonel₁ gave a speech. **He**₁ tried to get a certain friendly tone into his voice’ (Linna, 1954/1999, p. 144).
9. *Lammio₁ huusi Mielosta₂, ja tämä₂ tuli sisään lähetit kannoillaan.* ‘Lammio₁ called for Mielonen₂, and **he**₂ came in with the messengers on his heels’ (Linna, 1954/1999, p. 286).

As the surface forms suggest, the Finnish demonstrative *tämä* and the Estonian pronoun *tema* originated from the same demonstrative root (Kulonen et al., 2000; Larjavaara, 1986a, 1986b), which developed differently in Finnish and Estonian—it remained a demonstrative in Finnish and become a pronoun in Estonian.

TABLE 1
Finnish and Estonian Anaphoric Forms

<i>Variable</i>	<i>Finnish</i>	<i>Estonian</i>
Personal pronoun	<i>Hän</i> ‘she/he’	<i>Ta</i> (long form: <i>tema</i>) ‘she/he’
Demonstrative pronoun	<i>Tämä</i> ‘this’	<i>See</i> ‘this’

Unlike the pronoun *hän*, the anaphoric demonstrative *tämä* ‘this’ has been characterized as referring to characters in the background (Varteva, 1998). Sulka and Karjalainen (1992) noted that *tämä* is “used to indicate the last mentioned out of two or more possible referents” (pp. 282–283), and Saarimaa (1949) stated that it refers to a recently mentioned, non-subject referents (as in Example 9). Further support for these claims comes from a corpus study by Halmari (1994), showing that *hän* prefers subjects, and *tämä* often refers to objects (see also Kaiser, 2000).

However, as with Estonian *ta* and *see*, the referential properties of *hän* and *tämä* cannot be fully explained unless we distinguish the effects of the antecedent’s grammatical role from effects of its linear position (Kaiser, 2003; 2005; Kaiser & Trueswell, 2008; see also Järvi­kivi et al., 2005). However, in this article, I do not investigate effects of word order; the main focus is on sentences where subjects precede objects.

In sum, the Finnish and Estonian anaphoric paradigms resemble each other in that third-person human antecedents can be referred to with personal pronouns, as well as with demonstratives. However, in Finnish, there is only one form of the third-person pronoun (*hän*), which contrasts with the presence of both a long, emphatic form and a short, default form of the personal pronoun (*tema* and *ta*) in Estonian, as illustrated in Table 1.

Overview of Corpus Analyses

I use data from parallel corpora of Estonian and Finnish, with the specific aim of evaluating the salience and contrast approaches for the Estonian long form *tema*, and the more general aim of investigating whether different anaphoric forms can be ranked along a unified salience scale or whether a more multidimensional, form-specific account is needed. The first two analyses investigated the salience account. The first analysis focused on what form is used in Finnish texts when Estonian uses the long form *tema*, and what the antecedent’s grammatical role is. In the second analysis, I looked at what form is used in Estonian texts when Finnish uses the demonstrative *tämä*. The results fail to provide clear support for the salience approach: Neither the forms used in Finnish, nor the grammatical

role analysis, provide positive evidence for the idea that *tema* is used for lower-salience antecedents. In the final part of the article, I investigate the role of contrast, building on Pajusalu's (1995, 1997) observations. The data confirm that *tema* is indeed sensitive to contrast. By applying existing (non-pronoun related) work on contrast to Estonian, I show that the referential properties of *tema* can be explained by the felicity conditions on contrastive foci (in the sense of Rooth, 1992) and contrastive topics (in the sense of Büring, 2003).

WHAT FORM IS USED IN FINNISH WHEN ESTONIAN USES THE LONG FORM *TEMA*?

In this corpus study, I analyzed 50 occurrences of Estonian *tema* in prose text to see (a) what form they correlate with in Finnish and (b) what the grammatical role of the antecedent is.

If we find that Estonian *tema* shows up in Finnish as the demonstrative *tämä*, a full NP, or a name—but not the pronoun *hän*—this would fit straightforwardly with the salience account (i.e., the claim that the long form *tema* is used for lower-salience referents). Furthermore, independently of what form is used in Finnish, finding that *tema* refers primarily to antecedents in non-subject position would also support the salience account, given that (preverbal and agentive) subjects are widely regarded as being more salient than arguments in other positions. (As mentioned in the introduction, a range of factors has been found to influence salience. Because a full analysis of the various factors is beyond the scope of this article, I focus on grammatical role as a measure of referent salience, given that a large body of literature, in a range of languages, has found that entities in subject position are more salient/prominent than entities in non-subject position.)

On the other hand, according to the contrast account, there is no reason to expect *tema* to correlate with the Finnish demonstrative *tämä*. Existing research shows that *tämä* is used for lower-salience referents, not for contrastive referents. In fact, Finnish has no morphologically distinct referential form specialized for contrast: Presumably, the regular personal pronoun *hän* 'she/he' would be used, perhaps with a focal accent, as has been argued for English (e.g., de Hoop, 2003). Furthermore, the contrast account does not predict the antecedents of *tema* to have a bias toward a particular grammatical role, as antecedents in any position can be contrastive.

Corpus and Coding

In this study, 50 occurrences of *tema*, as well as their Finnish counterparts, from four novels (see Example 10) were analyzed. Two of the novels were

originally written in Finnish, and two in English. Given that neither the Finnish nor the English anaphoric paradigms (the two source languages of the translations) have a long versus short third-person pronoun distinction, a translator's decision to use *tema* rather than *ta* or *see/too* cannot be attributed to influence from the source language. This means that the material is well-suited for our purposes.

The analysis only included nominative occurrences of *tema* in subject position. The analysis was restricted in this manner because, as Pajusalu (1997) noted, in Estonian, the *taltema* variation is restricted for more oblique cases, but both forms can occur in subject position. Because I did not want to potentially bias the results by looking at structures that impose grammatical restrictions on which anaphoric forms can occur, I focused on the subject position, which permits the long and short personal pronouns, as well as demonstratives. In addition, I excluded occurrences of *tema* in noun–noun coordinations because *ta* is not possible in noun-level coordinations (see also Cardinaletti & Starke, 1999). I also excluded uses of *tema* modified by *ka* 'also' or 'too' or by the clitic *-ki/-gi* 'also', 'too', or 'as well' because these modifiers strongly prefer *tema* and rarely, if ever, occur with *ta* (see Tauli, 1983, p. 334). (It is worth noting that because these elements are often regarded as focus-sensitive, the strong preference for *tema* with these elements supports the contrast account; see also Sauerland, 2000.) Each occurrence of *tema* was coded for (a) the grammatical role of the most recent preceding mention of the antecedent in the Estonian version of the text and (b) what form was used in the Finnish version.

- 10a. Finnish original: Utrio, K. (1989). *Vendela*. Helsinki, Finland: Tammi.
Estonian translation: Utrio, K. (1996). *Vendela* (M. Jürima, Trans.).
Tallinn, Estonia: Sinisukk.
- 10b. Finnish original: Utrio, K. (1992). *Vaskilintu*. Helsinki, Finland: Tammi.
Estonian translation: Utrio, K. (1994). *Vasklind* (A. Lepp, Trans.).
Tallinn, Estonia: Sinisukk.
- 10c. English original: Rowling, J. K. (2000). *Harry Potter and the goblet of fire*. New York: Scholastic.
Finnish translation: Rowling, J. K. (2001). *Harry Potter ja liekehtivä pikari* (J. Kapari, Trans.). Helsinki, Finland: Tammi.
Estonian translation: Rowling, J. K. (2000). *Harry Potter ja tulepeeker* (K. Kaer & K. Kaer, Trans.). Tallinn, Estonia: Varrak.
- 10d. English original: Rowling, J. K. (1999). *Harry Potter and the prisoner of Azkaban*. New York: Scholastic.

Finnish translation: Rowling, J. K. (2001). *Harry Potter ja Azkabanin vanki* (J. Kapari, Trans.). Helsinki, Finland: Tammi.

Estonian translation: Rowling, J. K. (2000). *Harry Potter ja Azkabani vang* (K. Kaer & K. Kaer, Trans.). Tallinn, Estonia: Varrak.

Results and discussion. Let us first consider the results regarding the Finnish counterpart of *tema*, and then look at the grammatical role of the antecedent. As shown in Figure 1 and Table 2, when Estonian uses the emphatic form *tema*, the most common form used in the Finnish versions of the texts is the regular personal pronoun *hän* ‘she/he’ (80%), $\chi^2(4, N = 50) = 113.00, p < .001$. The overwhelming preference for *hän*, and the infrequent occurrence of the Finnish demonstrative *tämä* ‘this’ (2%) (as well as the low frequency of full NPs and other similar forms), shows that the referential properties of the long form *tema* do not resemble the referential properties of *tämä*. Thus, these findings do not provide support for the claim that *tema* is used for lower-salience referents.

Let us now turn to the data regarding the grammatical role of the antecedent. As Figure 2 and Table 3 show, the most frequent antecedent for *tema* is the preceding subject (38%; see Example 11), followed by direct and oblique objects (24%). The number of subject antecedents and object antecedents do not significantly differ from each other, $\chi^2(1, N = 31) = 1.581, p > .2$. Thus, *tema* does not exhibit a strong subject preference like the short pronoun *ta* does (e.g., see Kaiser & Hiietam, 2004; Kaiser & Vihman, in press); but, at the same time, the distribution of subject and object antecedents indicates that *tema* cannot be characterized as being strongly anti-subject.

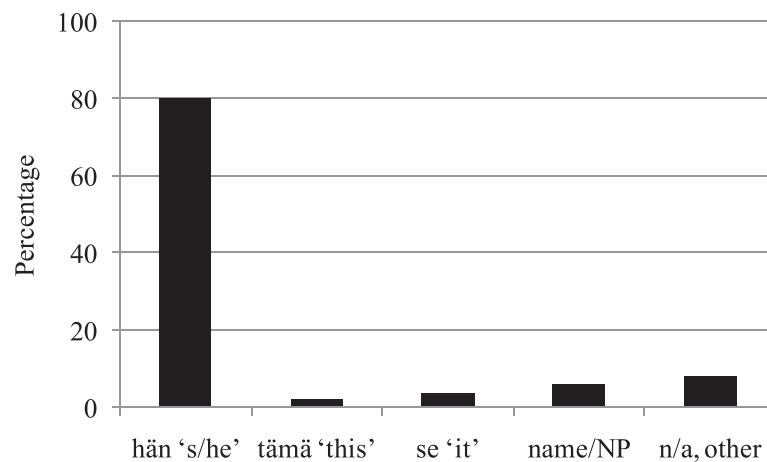


FIGURE 1 Counterparts of Estonian long pronoun *tema* in the Finnish versions of the texts. *Note.* NP = noun phrase.

TABLE 2
Counterparts of Estonian *Tema* in the Finnish Versions of the Texts

<i>Variable</i>	<i>%</i>	<i>Raw Counts</i>
<i>Hän</i> 'she/he'	80	40/50
<i>Tämä</i> 'this'	2	1/50
<i>Se</i> 'it'	4	2/50
Name/noun phrase	6	3/50
NA/other	8	4/50

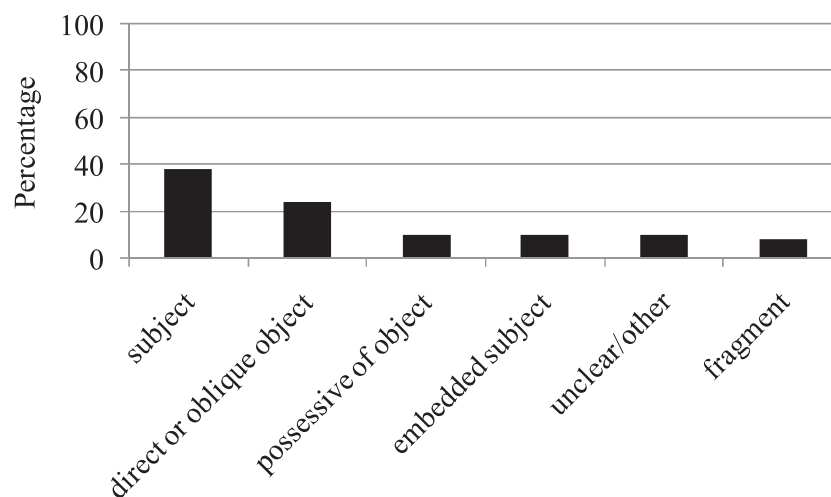


FIGURE 2 Grammatical role of the antecedent of *tema* in the Estonian texts.

TABLE 3
Grammatical Role of the Antecedent of *Tema* in the Estonian Texts

<i>Variable</i>	<i>%</i>	<i>Raw Counts</i>
Subject	38	19/50
Direct or oblique object	24	12/50
Possessive of object/oblique	10	5/50
Embedded subject	10	5/50
Unclear/other	10	5/50
Fragment ^a	8	4/50

^aFragment means that the antecedent occurred in an exclamation or some other kind of "sentence fragment" that did not constitute a complete sentence.

11. [Context: Lupin’s rights to attend the Hogwarts school are in danger because of other parents’ possible concerns about him—being a werewolf—and being near their children]

*Aga siis sai **Dumbledore**₁ direktoriks. . . . **Ta**₁ ütles, et senikaua kui me võtame tarvitusele teatud abinõud, ei näe **tema**₁ mingit põhjust, miks ma ei peaks koolu tulema.* ‘But then **Dumbledore**₁ became Headmaster. . . . **He**₁ said that as long as we took certain precautions, **he**₁ saw no reason why I shouldn’t come to school’ (Rowling, 2000, p. 305).

In the original English text, the structure is slightly different: “. . . He said that as long as we took certain precautions, there was no reason I shouldn’t come to school.” In Example 11, I provide the English translation of the Estonian text.

In sum, based on the corpus counts of the grammatical role of the antecedent of *tema* and the Finnish counterparts of *tema*, we can conclude that (a) the long pronoun *tema* does not have similar referential properties as the Finnish demonstrative *tämä*, as shown by the fact that *tema* is usually realized as *hän* ‘she/he’ in Finnish; and (b) *tema* does not have a clear preference for non-subjects over subjects. Thus, these results fail to provide straightforward evidence in favor of the salience account, which posits that long forms are used for lower-salience referents. Although these results are not necessarily clear evidence against all variants of the salience account, and cannot be used to dismiss that account outright (I return to this issue in the Conclusions and Future Work Section), the finding that two different ways of tapping into potential salience effects failed to find any evidence in its favor does seem to cast doubt on the salience account.

What Form Is Used in Estonian When Finnish Uses the Demonstrative *Tämä*?

As an additional measure, I analyzed what form is used in Estonian in a situation where Finnish uses the demonstrative *tämä* ‘this’. Given that Finnish *tämä* is known to be used for lower-salience antecedents, we can investigate what form Estonian uses to refer to such antecedents. I analyzed 74 tokens of *tämä* (when used to refer to human antecedents) in subject position, from four Finnish novels translated into Estonian by four different translators (see Example 12):

- 12a. Finnish original: Paasilinna, A. (1975). *Jäniksen Vuosi*. Helsinki, Finland: Weilin & Göös.
Estonian translation: Paasilinna, A. (1995). *Jänese Aasta* (S. Kiin, Trans.). Tallinn, Estonia: Perioodika.

- 12b. Finnish original: Paasilinna, A. (1998). *Ulvova Mylläri*. Helsinki, Finland: Werner Söderström Osakeyhtiö. (Original work published 1981)
Estonian translation: Paasilinna, A. (1987). *Ulguv Mölder* (S. Ruutsoo, Trans.). Tallinn, Estonia: Perioodika.
- 12c. Finnish original: Paasilinna, A. (1983). *Hirtettyjen Kettujen Metsä*. Helsinki, Finland: Werner Söderström Osakeyhtiö.
Estonian translation: Paasilinna, A. (1997). *Poodud Rebaste Mets* (K. Kokk, Trans.). Tallinn, Estonia: Monokkel.
- 12d. Finnish original: Utrio, K. (1989). *Vendela*. Helsinki, Finland: Tammi.
Estonian translation: Utrio, K. (1996). *Vendela* (M. Jürima, Trans.). Tallinn, Estonia: Sinisukk.

These novels overlap only partially with those used for the earlier corpus study. This is because I wanted to use all-Finnish source texts for this analysis to ensure that the Estonian translators were working from Finnish texts because the aim of this analysis was to see how they translate the Finnish demonstrative *tämä* 'this' into Estonian.

Results and discussion. The results show that, most often, the Finnish demonstrative *tämä* is translated into Estonian as the demonstrative *see/too* (54%), as shown in Figure 3 and Table 4. The demonstrative *see/too* is used significantly more often than the other options, including names and full NPs, $\chi^2(1, N = 56) = 10.286, p < .01$; and the short pronoun *ta* 'she/he', $\chi^2(1, N =$

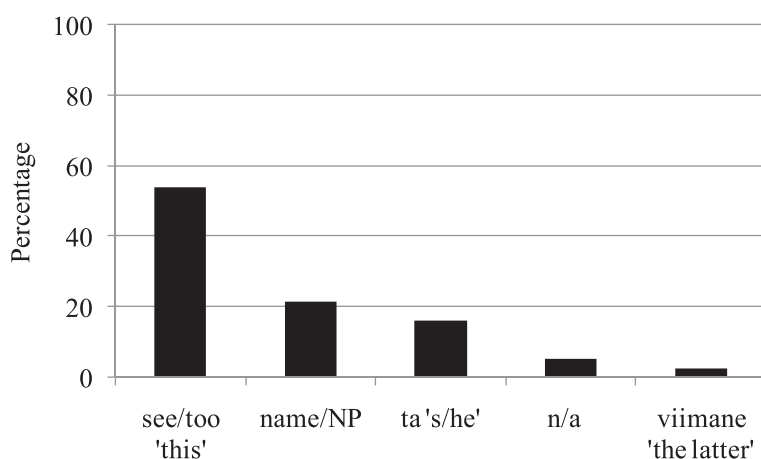


FIGURE 3 Counterparts of the Finnish demonstrative *tämä* in the Estonian versions of the texts. *Note.* There were a few cases where the Estonian translation was structurally different, so there was no referent in subject position to use for comparison. NP = noun phrase.

TABLE 4
Counterparts of the Finnish Demonstrative *Tämä*
in the Estonian Versions of the Texts

<i>Variable</i>	<i>%</i>	<i>Raw Counts</i>
<i>See/too</i> 'this'	54.1	40/74
Name/noun phrase	21.6	16/74
<i>Ta</i> 'she/he'	16.2	12/74
NA	5.4	4/74
<i>Viimane</i> 'the latter'	2.7	2/74

52) = 15.077, $p < .001$.¹ An example of *tämä* translated as *see* is provided in Example 13. Crucially, no occurrences of Finnish *tämä* were translated into Estonian using the long form *tema*, despite the surface similarity and shared origins of these words. (Grammatical roles of the antecedents of the Finnish demonstrative *tämä* and the Estonian demonstratives *see/too* were also analyzed. As expected, the demonstratives in both languages exhibit a strong preference for non-subject antecedents; for details, see Kaiser, 2003).

13. *Vatanen (1) sanoi taksimiehelle (2), että tämä (2) ajaisi nopeasti Kuopion keskussairaalan poliklinikalle.* (Finnish, Paasilinna, 1975, p. 82)

Vatanen (1) ütles taksojuhile (2), et see (2) sõidaks kiiresti Kuopio keskihaigla polikliinikusse (Estonian, Paasilinna, 1995, p. 52)

'Vatanen (1) told the taxi driver (2) that he (2) should drive quickly to the Kuopio central hospital polyclinic.'

In sum, these results fit with the outcome of the first analysis in that they do not provide any direct support for the salience account of *tema*. The finding that Finnish *tämä* tends to be translated into Estonian as *see/too* supports the claim that, in Estonian, these demonstratives are used for lower-salience antecedents—much like *tämä*. In addition, the frequent use of full NPs or names is not unexpected because these forms are used for referents that are not currently at the center of attention (see Ariel, 1990; Givón, 1983; Gundel et al., 1993). There are no cases of Finnish *tämä* being translated into Estonian as *tema*, which fits with the first analysis.

¹If this is broken down into the two variants, *see* and *too*, it becomes clear that *see* is more frequent (36% of all occurrences of *tämä*), with *too* used less often (18%). All but one of the occurrences of *too* were from the work of translator Kādi Kokk and can, thus, probably be attributed to the dialectal background of that particular translator.

Contrast

In the preceding sections, we saw that the antecedents of the long form *tema* are not associated with low-salience grammatical roles, and that the long form *tema* does not map onto the Finnish low-salience anaphor *tämä* ‘this’. In this section, I explore in more detail Pajusalu’s (1997) idea that *tema* is used to refer to entities that are being compared or opposed to some other referents.

Different researchers use the term *contrast* in different ways (e.g., É. Kiss, 1998; Rooth, 1992; Vallduví & Vilkuna, 1998; Zimmermann, 2007; see also Molnár, 2006, for an overview), but it is generally agreed that, at least on an intuitive level, contrast has to do with alternatives—in particular, contextually salient alternatives. In the next section, to formalize this notion of alternatives, I first discuss *tema* from the perspective of alternative semantics (e.g., Rooth, 1985, 1992). Then, using examples from Estonian, I take a closer look at what kind of contrast is relevant for the use of *tema*. I suggest that the referential properties of *tema* can be explained if we combine its pronominal nature with the felicity conditions for contrastive focus and contrastive topics (discussed later).

Focus in Alternative Semantics

According to alternative semantics, a sentence that contains focused material has both an ordinary semantic value $[[.]]^o$ and a focus semantic value $[[.]]^f$. The focus semantic value results from replacing the focused expression with its alternatives and, thus, the focus semantic value of a sentence is a set of propositions. This is shown in Example 14 for a sentence where the object “Mari” is focused (as indicated by the brackets and the “F” subscript). Let us call the set of propositions in Example 14b the *focus* set of alternatives for Example 14a:

- 14a. Liisa tickled [MARI]_F.
 14b. $[[\text{Liisa tickled MARI}]]^F = \{\text{Liisa tickled Mari, Liisa tickled Anna, Liisa tickled Matt}\}.$

Rooth (1992) discussed the distinction between exhaustive focus and contrastive focus; the present discussion focuses on contrastive focus. Following Rooth (1992; see also Romero & Han, 2003), I assume that contrastive focus is felicitous when one of the members of the focus set of alternatives evoked by the focus-containing sentence was already mentioned or is otherwise salient in preceding discourse. If the alternatives are not provided by preceding discourse, they presumably need to be accommodated in order to make focusing felicitous. Thus, a sentence such as “Liisa tickled MARI,” with contrastive focus on Mari,

triggers the interpretation that Liisa tickled Mari, rather than one of the other people she could potentially have tickled, such as Anna or Matt (the alternatives). Essentially, the contrastively focused entity is being compared or opposed to the other members of the alternative set.

Estonian Data

Using the notion of contrastive focus as a starting point, in this section I discuss the kinds of contrastive configuration that were observed in the corpus. (In the subsequent discussion, I assume that *tema* is focused and evokes a set of alternatives. This is compatible with existing claims that *tema* can be [but does not need to be] stressed because focus is usually assumed to be associated with prosodic stress/accent. Later, I discuss the question of stress/accent more.)

In Example 15, *tema* refers to a woman called Vendela in a context where Father Henrik has just claimed that he will take care of the knight, Sir Hartman. The contextually salient focus set of alternatives evoked by the sentence with *tema* is in Example 15b. One of the members of this set was already mentioned (“Father Henrik will take care of the knight”) and, therefore, contrastive use of *tema* in Example 15a is felicitous. Thus, here *tema* is used when *denying a previously made assertion* (on corrections, see Umbach, 2004). Furthermore, because *tema* is a pronoun, it must refer to an antecedent that is sufficiently salient in preceding discourse. If we combine this requirement with the felicity conditions on contrastive focus, we can capture this type of *tema* usage.

- 15a. [Context: Father Henrik wants to come along to take care of Sir Hartman, who is seriously ill. The head of Sir Hartman’s men explains the following to him]

See ... sõltub täielikult sellest, kas Domina Vendela lubab sul kaasa tulla või mitte. Domina Vendela on ravitseja. Tema ravib rüütlit ... ‘It ... depends entirely on whether Domina Vendela allows you to come along or not. **Domina Vendela** is a healer. **She** will take care of the knight ...’ (Utrio, 1996, p. 94).

- 15b. {Father Henrik will take care of the knight, Vendela will take care of the knight}.

In addition to tokens where *tema* is used when denying a preceding assertion, the corpus contains tokens where *tema* is used when making *contrasting assertions about different entities*. For example, in Example 16a, *tema* is used to refer to Sir Hartman in the sentence, “He could not read,” which follows mention of the fact that Vendela can read:

- 16a. [Context: Vendela has just told Sir Hartman that she can read and that she even owns a book, which was quite a rare possession in Finland in the year 1371]

Rüütel Hartman mõtiskles selle üle, lebedes mõnusalt laas voodis. Tema ei osanud lugeda, selleks polnud mingit vajadust—lugemine oli pastorite osa. ‘**Sir Hartman** thought about this, resting comfortably in the wide bed. **He** could not read, there was no need for it—reading was for pastors’ (Utrio, 1996, p. 107).

- 16b. {Hartman cannot read, Hartman can read, Vendela cannot read, Vendela can read}.

If we assume that both *tema* and negation are focused, the set of alternatives evoked by the sentence with *tema* is as shown in Example 16b.² One of the members of this focus-evoked set was already mentioned in preceding discourse—namely, “Vendela can read”—and, thus, use of contrastively focused *tema* is felicitous. Thus, this use of *tema*—which involves making a contrasting assertion about another entity—also follows from (a) the requirement that *tema* have a salient antecedent and (b) the felicity conditions on contrastive focus.

In the examples considered so far, one of the members of the focus set of alternatives evoked by the focus-containing sentence was already explicitly mentioned in preceding discourse. However, this is not always the case, as exemplified in Example 17:

- 17a. [Context: A new woman is arriving at the village. Everyone has gathered to look at her, including another woman called Hailvi]

Naine oli ümmarguse näo ja väikese suuga. Tema vammus oli valmistatud kirkalt läikivast kangast; Rodarve Hailvi ütles, et see on siid, ja tema juba teadis, sest Rodair Rodmundsson oli toonud talle siidrieti Gaardarikest. ‘The woman had a round face and a small mouth. Her coat was made from brightly sparkling material; **Hailvi of Rodarve** said that is silk, and **she** indeed knew, because Rodair Rodmundsson had brought her silk from Gardarike’ (Utrio, 1994, p. 98).

²See Han and Romero (2001, p. 268) for a discussion of a sentence with a focused subject and focused polarity–negation. I assume examples like 16 and 17 contain two focused elements (*tema* and polarity). Such an approach builds on work by Han and Romero and others. However, if we adopt an analysis along the lines of Büring (2003), then, depending on the context, examples with *tema* in subject position and focused polarity could arguably be analyzed as involving one focus and one contrastive topic (CT), rather than two foci. Thank you to Daniel Büring for helpful discussion in this area. However, whether examples like (16) and (17) are best analyzed as involving two foci or a contrastive topic and focus does not have consequences for the claims regarding *tema* made in this article.

- 17b. {Hailvi knows that the material is silk, Hailvi does not know that the material is silk, Person 1 knows that the material is silk, Person 1 does not know that the material is silk, Person 2 knows that the material is silk, Person 2 does not know that the material is silk ... }.

In this example, *tema* triggers the inference that Hailvi, unlike the other people watching the new woman's arrival, knew that the material of her coat was silk. I assume that, similar to Example 16, not only *tema*, but also the polarity of the sentence, is focused (note the presence of *juba* 'indeed'), resulting in the focus set of alternatives shown in Example 17b. However, this focus set of alternatives was *not explicitly mentioned in the preceding discourse*—but it is clearly plausible and can be accommodated. Examples requiring this kind of plausible accommodation are frequent in the corpus, and show that *tema* is capable of evoking a focus set of alternatives. In other words, it is not limited to contrasting with already-mentioned alternatives.

Focus and Contrastive Topics

So far, we have considered examples like 15, where *tema* is used to deny a preceding assertion; and examples like 16 and 17, with focus on *tema* and on negation, where the negated form of the sentence with *tema* has been claimed to hold of another referent. *Tema* also occurs in sentences containing another focused element. Consider Example 18. Here, some of the characters in *Harry Potter* are watching famous people walk by their campsite, and Mr. Weasley is providing information about each person's job. Another similar example Example 7 from Pajusalu (1997). In Example 7, *tema* is also used in a list context, where the role of the first dancer had been previously mentioned and the speaker is now specifying the role of the second dancer. A third example of this type is in Example 19. Here, *tema* is used to refer to the bailiff, Harittu, in a context where Iliana is questioning Vendela about different people and their backgrounds. (The second occurrence of *tema* in this example is in the genitive case and, thus, not analyzed, as explained earlier):

18. [Context: Mr. Weasley comments on passers-by, as Harry and Hermione do not know them]

See oli Cuthbert Mockridge, Härjapõlvlastega Sidemete Arendamise Ameti juhataja ... sealt tuleb Gilbert Wimble, tema on Eksperimentaalloitsude Komitees, tal on on need sarved juba päris tükk aega olnud ... Tere, Arnie ... Arnold Peasegood, tema on mälukustutaja—teate, völlumisäparaduste heastamise rühmast ... (Rowling, 2000, p. 75).
 'That was Cuthbert Mockridge, Head of the Goblin Liason Office ...

Here comes Gilbert Wimple, **he**'s with the Committee on Experimental Charms, he's had those horns for a while now ... Hello Arnie ... Arnold Peasegood, **he**'s an Obviator—member of the Accidental Magic Reverse squad, you know ...' (Rowling, 2000, p. 86).

19. [Context: Iliana has been asking Vendela what different people are like. She has asked about the priest, Father Henrik, and about another woman, Mielitty. Then she asks about the bailiff]

*“Kuid foogt?” jätkas Iliana. “Missugune tema on? Ja tema nainen?”
“Harittu on Tuurilas kauan olnud,” sositas Vendela hirmunult. (Utrio, 1996, p. 155)*

“And what about the **bailiff**?” continued Iliana. “What is **he** like? And his wife?”

“Harittu has been at Tuurila for a long time,” Vendela whispered fearfully.

If we assume that *tema* is focused, then these sentences contain two foci (e.g., *x* has job *y* [see Example 18], *x* dances the role of *y* [see Example 7], and *x* has personality trait *y* [see Example 19]). In fact, these examples fit well with a class of English examples discussed by Jackendoff (1972) and Büring (2003). Jackendoff noted that, in English, a constituent that is prosodically focused can be accented in at least two different ways, which Büring labeled as “focus” and “contrastive topic.” For example, in Example 20, there are two accented elements, “Fred” and “beans,” realized with different accents (what Jackendoff, 1972, called the “B” accent and the “A” accent, respectively):

20a. Well, what about FRED? What did HE eat?

20b. [FRED]_{CT} ate the [BEANS]_F.

The question in Example 20a can felicitously be uttered in a context where we are discussing what Fred ate, what Tina ate, what Mary ate, and so on. In other words, the question in Example 20a is a sub-question within a larger question (“Who ate what?”; see also Roberts, 1996). In the answer in Example 20b, Fred is a member of a salient set under discussion (the set of people who ate something) and contrasts with the other members in terms of what he ate. Büring (2003) called FRED the contrastive topic (subscript CT) because Fred was also mentioned in the question and is old information. Büring referred to BEANS as the focus (subscript F), the new information that answers the question for each person. (Which parts of a sentence are construed as CT and F depends on the context. If someone asks, “Well, what about the BEANS? Who ate THEM?,” in addition to questions about who ate various other foods, the answer could be,

“FRED ate the BEANS,” with FRED as F, and BEANS as CT; see Jackendoff, 1972, p. 261; see also Buring, 2003).

The distinction between contrastive topic and focus can be applied to the Estonian examples in Examples 18, 7, and 19. As we are focusing on written data, I leave aside, for now, questions regarding the specific prosodic realization of *tema*. However, from a semantic perspective, in these examples, *tema* can be regarded as a contrastive topic, and the second focus is the new information that answers the implicit question, as exemplified in Example 21 for Example 7 and one of the occurrences of *tema* in Example 18:

- 21a. What about the second dancer? What role does she dance?
 [*tema*]_{CT} *tantsib* [*seltsidaami*]_F. ‘She dances (the role of the) lady-in-waiting.’
- 21b. What about Arnold Peasegood? What is his position?
 [*tema*]_{CT} *on* [*mälukustutaja*]_F. ‘He’s an Obviator.’

These corpus patterns are also supported by native speaker intuitions. In Example 22 (parallel to Example 20), native speakers judge *tema* to sound more felicitous than *ta*. However, in a situation where the referent does not contrast with anyone, the most felicitous form is *ta* (see Example 23):

- 22a. *Ma tean, et Liisa kõdistas Marit. Kuid keda kõdistas Anna?* ‘I know that Liisa tickled Mari. But who did Anna tickle?’
- 22b. [*#Ta/Tema*]_{CT} *kõdistas* [*Tiinat*]_F. ‘She tickled Tiina.’
- 23a. *Ma tean, et Anna müksas Marit. Ja mis peale seda sai?* ‘I know that Anna pushed Mari. And what happened after that?’
- 22b. [*Ta/#Tema*] *kõdistas* *Tiinat*. ‘She tickled Tiina.’

In sum, a detailed analysis of the corpus examples showed that use of *tema* is indeed sensitive to contrast, lending support to Pajusalu’s (1997) observation. It appears that in the prose texts analyzed in this article, at least in the syntactic positions that I focused on (where *ta* and *tema* are both grammatically possible), the occurrence of *tema* is sensitive to the presence of a contrastive focus or a contrastive topic, as described earlier. These findings contribute to our understanding of what type of contrast is relevant for *tema*: I suggest that the referential properties of Estonian long pronouns can be derived from (a) the pronoun-related requirement that *tema* have a sufficiently salient antecedent and (b) the felicity conditions on contrastive focus (in the sense of Rooth, 1992) and CT (in the sense of Buring, 2003).

It is worth pointing out that, depending on one’s theory of focus and contrast, it might be possible to provide a unified account of *tema*’s referential properties in terms of contrastive topics in Buring’s (2003) sense. This is a promising

direction for future work, and would benefit from larger-scale corpus studies, as well as more information about the intonational contour/pitch accent on *tema* in different contexts.

CONCLUSION AND FUTURE WORK

In this article, by using a corpus of naturally occurring examples of the Estonian third-person pronoun *tema*, I investigated its referential properties to find out whether the long form *tema* (as opposed to the short form *ta*) is used to mention referents that are not highly salient (salience account), or whether it is used for entities that are mentioned contrastively (contrast account). The corpus data indicate that the long form *tema* does not have a preference for non-subject antecedents, and that Estonian *tema* and Finnish *tämä* ‘this’ do not have the same referential properties. Thus, as a whole, the cross-linguistic corpus results failed to provide clear evidence in favor of the salience account. Instead, the corpus data support Pajusalu’s (1997) observation that contrast guides use of the Estonian long pronoun in structural positions where both the long and the short forms are possible. By analyzing the corpus examples in terms of alternative semantics (Rooth, 1992) and Büring’s (2003) definition of contrastive topic, I took steps toward a more explicit account of what type of contrast is relevant for use of *tema*.

However, despite the lack of any positive evidence for the salience account, the patterns observed in the first two corpus analyses do not justify an outright rejection of the salience account for *tema* for two main reasons. First, so far we have been working with the assumption that if the long form *tema* refers to lower-salience antecedents than the short form *ta*, it should be realized in Finnish with a form that is ranked below the pronoun *hän* on the salience scale (e.g., the demonstrative *tämä*, a full NP, etc.). However, what if both the Estonian short form *ta* and the long form *tema* map onto the Finnish pronoun *hän* on the salience scale? In this case, the Finnish demonstrative *tämä* would be ranked below both *ta* and *tema*, resulting in a situation where (a) *tema* is used for less salient antecedents than *ta* (as the salience account claims), but (b) *tema* and *ta* nevertheless both show up as *hän* in Finnish.

Second, recall that grammatical role is not the only factor that has been claimed to influence referent salience. Thus, one might ask whether the long form *tema* is sensitive not to grammatical role but to another salience-influencing factor (such as, say, implicit causality effects due to verb semantics; e.g., Garvey & Caramazza, 1974). (Presumably, the Finnish pronoun *hän* would have to be sensitive to this same factor, given that *tema* maps onto *hän* in the parallel corpora.) This would be a very interesting finding. Combined with Kaiser and Vihman’s (in press) finding that the Estonian short pronoun *ta* is primarily

sensitive to subjecthood/agentivity, it would fit nicely with the form-specific multiple-constraint approach, as we might have a situation where one form is most sensitive to grammatical role, whereas another form is primarily sensitive to another factor, or perhaps a weighted combination of grammatical role and another factor. Thus, from a broader theoretical perspective, both the contrast account (i.e., *tema* is used for referents that contrast with other entities in the discourse model; see later discussion) and this “alternative salience factor” account seem to be compatible with the form-specific multiple-constraints approach, as they highlight the varying sensitivities of different forms.

In sum, the patterns observed in the first two corpus analyses are not necessarily incompatible with a salience account. However, given that multiple ways of tapping into potential salience effects failed to find positive evidence in favor of the salience account, it seems that other possibilities, such as the role of contrast, merit consideration. The absence of clear evidence for the salience account, combined with the finding that *tema* patterns with contrastive topics and contrastive foci, indicates that, although some version of a salience account may turn out to be compatible with the data, it is not sufficient on its own.

The finding that use of the long pronoun *tema* is sensitive to the presence of contrast, combined with earlier research showing that the demonstrative *see* and the short pronoun *ta* appear more sensitive to grammatical role, linear order, and salience (Kaiser & Hiietam, 2004; Kaiser & Vihman, in press), fits well with the form-specific multiple-constraints account (e.g., Kaiser & Trueswell, 2008). As a whole, the findings for Estonian long and short pronouns provide further cross-linguistic evidence for the form-specific approach, and show that even two closely related forms can exhibit asymmetrical sensitivities. The findings regarding contrast show that referential forms can differ not only in their sensitivity to antecedent properties that are explicitly linguistically encoded (e.g., word order, grammatical role; Kaiser & Trueswell, 2008; Kaiser & Vihman, in press), but also in how sensitive they are to a more abstract, discourse-dependent notion such as contrast.

An important and under-investigated question concerns the prosodic realization of *tema*. Researchers have suggested that *tema* can, but does not need to, be stressed. What has not yet been investigated in detail is what kind of accent *tema* has under different discourse conditions—something that could be investigated with a spoken corpus or by means of a production experiment. In general, psycholinguistic experiments offer a promising avenue for future research, as they would allow us to test comprehension, as well as production, in specific discourse contexts. In addition, a larger corpus of *tema* should be collected to see if the patterns I observed also occur on a larger scale and in different types of corpora (spoken, as well as written).

In closing, it is worth emphasizing that the conclusion that use of *tema* is sensitive to contrast does not mean that the concept of salience should be

abandoned. For example, it seems reasonable to posit that the choice of reduced forms (e.g., “he”) versus fuller anaphoric forms (e.g., “the little Finnish boy that I told you about yesterday”) depends on how prominent the intended referent is in the language users’ mental models of the discourse. As a whole, the results presented here—as well as the findings of Kaiser (2003, 2005), Brown-Schmidt et al. (2005), and Kaiser and Trueswell (2004, 2008)—are best regarded as evidence that reference resolution is a multifaceted process, with anaphoric forms exhibiting different levels of sensitivity to different properties.

ACKNOWLEDGMENTS

I thank Jeanette Gundel, Peter Bosch, Daniel Büring, Christine Gunlogson, Renate Pajusalu, Ellen Prince, Maribel Romero, and Carla Umbach for suggestions and feedback at various stages of this project, as well as Renate Pajusalu, Katrin Hiietam, Helena Seli, Mari Allik, and Virve Vihman for help with Estonian. I also thank three anonymous reviewers for useful feedback and suggestions. Parts of this article are based on aspects of chapter 4 of my dissertation (Kaiser, 2003). Earlier versions of some of this research were presented at the 28th Penn Linguistics Colloquium (with a proceedings article in the Penn Working Papers in Linguistics), the Texas Linguistic Society Meeting 8 (“Issues at the Semantics–Pragmatics Interface”), and the workshop on Anaphoric Uses of Demonstrative Expressions (at the 29th annual meeting of the Deutsche Gesellschaft für Sprachwissenschaft).

REFERENCES

- Alonso-Ovalle, L., Clifton, C., Fernández-Solera, S., & Frazier, L. (2002). Null vs. overt pronouns and the topic-focus articulation in Spanish. *Journal of Italian Linguistics*, 14, 151–169.
- Ariel, M. (1990). *Accessing NP antecedents*. London: Routledge/Croom Helm.
- Ariel, M. (2001). Accessibility theory: An overview. In T. Sanders, J. Schilperoord, & W. Spooren (Eds.), *Text representation, linguistic and psycholinguistic aspects* (pp. 29–87). Amsterdam: Benjamins.
- Arnold, J. E. (1998). *Reference form and discourse patterns*. Unpublished doctoral dissertation, Stanford University, Stanford.
- Bosch, P., Rozario, T., & Zhao, Y. (2003). Demonstrative pronouns and personal pronouns. German *der* vs. *er*. In *Proceedings of the EACL 2003 workshop on the Computational Treatment of Anaphora, Association for Computational Linguistics, 10th Meeting of the European Chapter* (pp. 61–68).
- Brennan, S. E., Friedman, M. A., & Pollard, C. J. (1987). A centering approach to pronouns. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics* (pp. 155–162). Stanford, CA: Association for Computational Linguistics.
- Brown-Schmidt, S., Byron, D. K., & Tanenhaus, M. K. (2005). Beyond salience: Interpretation of personal and demonstrative pronouns. *Journal of Memory and Language*, 53, 292–313.

- Büring, D. (2003). On D-trees, beans and B-accents. *Linguistics & Philosophy*, 26, 511–545.
- Cardinaletti, A., & Starke, M. (1999). The typology of structural deficiency: A case study of the three classes of pronouns. In H. van Riemsdijk (Ed.), *Clitics in the languages of Europe* (pp. 145–233). Berlin: Mouton.
- Carminati, M. N. (2002). *The processing of Italian subject pronouns*. Unpublished doctoral dissertation, University of Massachusetts, Amherst.
- Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 25–55). New York: Academic.
- Chambers, C., & Smyth, R. (1998). Structural parallelism and discourse coherence: A test of Centering Theory. *Journal of Memory and Language*, 39, 593–608.
- Crawley, R., & Stevenson, R. (1990). Reference in single sentences and in texts. *Journal of Psycholinguistic Research*, 19, 191–210.
- de Hoop, H. (2003). On the interpretation of stressed pronouns. In M. Weisgerber (Ed.), *Proceedings of the 7th Sinn und Bedeutung conference* (pp. 159–172). Germany: Universität Konstanz.
- É. Kiss, K. (1998). Identificational focus versus information focus. *Language*, 74, 245–273.
- Erelt, M., Kasik, R., Metslang, H., Rajandi, H., Ross, K., Saari, H., et al. (1993). *Eesti keele grammatika II Süntaks* [Estonian grammar II. Syntax]. Tallinn, Estonia: Eesti Teaduste Akadeemia Eesti Keele Instituut [Estonian Science Academy, Estonian Language Institute].
- Gahl, S., & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*, 80, 748–775.
- Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry*, 5, 459–464.
- Gernsbacher, M. A., & Hargreaves, D. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, 27, 699–717.
- Givón, T. (1983). *Topic continuity in discourse*. Amsterdam: Benjamins.
- Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69, 274–307.
- Haeseryn, W., Romijn, K., Geerts, G., de Rooij, J., & van den Toorn, M. C. (1997). *Algemene Nederlandse Spraakkunst, Band 1* [General Dutch grammar, Vol. 1]. Groningen: Martinus Nijhoff.
- Halmari, H. (1994). On accessibility and coreference. *Nordic Journal of Linguistics*, 17, 35–59.
- Han, C.-h., & Romero, M. (2001). Negation, focus and alternative questions. In K. Megerdooian & L. A. Bar-el (Eds.), *Proceedings of the 20th West Coast Conference on Formal Linguistics* (pp. 262–275). Somerville, MA: Cascadilla.
- Jackendoff, R. S. (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
- Järvikivi, J., van Gompel, R. P. G., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution: Contrasting the first-mention and subject preference accounts. *Psychological Science*, 16, 260–264.
- Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–254). Amsterdam: Benjamins.
- Kaiser, E. (2000). Pronouns and demonstratives in Finnish: Indicators of referent salience. In P. Baker, A. Hardie, T. McEnery, & A. Siewierska (Eds.), *Proceedings of the Discourse Anaphora and Anaphor Resolution Conference* (pp. 20–27). Lancaster, UK: Lancaster University.
- Kaiser, E. (2003). *The quest for a referent: A crosslinguistic look at referent resolution*. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia, PA.
- Kaiser, E. (2005). When salience isn't enough: Pronouns, demonstratives and the quest for an antecedent. In R. Laury (Ed.), *Minimal reference in Finnic: The use and interpretation of pronouns and zero in Finnish and Estonian discourse* (pp. 135–162). Helsinki, Finland: Suomalaisen Kirjallisuuden Seura.
- Kaiser, E., & Hiitam, K. (2004). A comparison of the referential properties of third person pronouns in Finnish and Estonian. *Nordlyd*, 31, 654–667.

- Kaiser, E., & Trueswell, J. C. (2004). The referential properties of Dutch pronouns and demonstratives: Is salience enough? In C. C. Meier & M. Weisgerber (Eds.), *Proceedings of the 8th Sinn und Bedeutung Conference* (pp. 137–149). Working Paper #177, Linguistics Department, University of Constance. Constance, Germany: Universität Konstanz.
- Kaiser, E., & Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference. *Language and Cognitive Processes*, 23, 709–748.
- Kaiser, E., & Vihman, V.-A. (in press). On the referential properties of Estonian pronouns and demonstratives. *Proceedings of the 22nd Scandinavian conference of linguistics*, Aalborg University, Denmark.
- Kaiser, E., Runner, J. T., Sussman, R. S. & Tanenhaus, M. K. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition*, 112, 55–80.
- Kalliokoski, J. (1991). Emphathy as motivation for style shifting in narrative. In J. Verschueren (Ed.), *Levels of linguistic adaptation* (Vol. 2, pp. 147–161). Amsterdam: Benjamins.
- Kameyama, M. (1999). Stressed and unstressed pronouns: Complementary preferences. In P. Bosch & R. van der Sandt (Eds.), *Focus. Linguistic, cognitive and computational perspectives* (pp. 306–321). UK: Cambridge University Press.
- Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI.
- Kehler, A., Kertz, L., Rohde, H., & Elman, J. (2008). Coherence and coreference revisited. *Journal of Semantics*, 25, 1–44.
- Kibrik, A. (1996). Anaphora in Russian narrative discourse: A cognitive calculative account. In B. Fox (Ed.), *Studies in anaphora* (pp. 255–304). Amsterdam: Benjamins.
- Kulonen, U.-M., et al. (Eds.) (2000). *Suomen sanojen alkuperä* [Origins of Finnish words; Etymological dictionary]. Helsinki, Finland: Suomalaisen Kirjallisuuden Seura.
- Larjavaara, M. (1986a). Demonstratiivit ja kielen kehitys [Demonstratives and the development of language]. *Virittäjä*, 90, 306–312.
- Larjavaara, M. (1986b). *Itämerensuomen demonstratiivit I* [Balto-Finnic demonstratives I]. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Linna, V. (1999). *Tuntematon sotilas* [Unknown soldier]. Helsinki: Werner Söderström Osakeyhtiö. (Original work published 1954)
- Luján, M. (1985). Binding properties of overt pronouns in null pronominal languages. In W. Eilfont, P. Kroeler, & K. Peterson (Eds.), *Proceedings of the 21st annual meeting of the Chicago Linguistic Society* (pp. 435–453). Chicago, IL: Chicago Linguistic Society.
- Molnár, V. (2006). On different kinds of contrast. In V. Molnár & W. S. Winkler (Eds.), *Architecture of focus* (pp. 197–233). Berlin: de Gruyter.
- Pajusalu, R. (1995). Pronominit *see*, *tema* ja *ta* viron puhekielessä [Pronouns *see*, *tema* and *ta* in spoken Estonian]. *Sananjalka*, 37, 81–93.
- Pajusalu, R. (1996). Pronoun systems of common Estonian and Estonian dialects in a contrastive perspective. In *Estonian: Typological studies* (Vol. 4, pp. 19–164), Tartu, Estonia: University of Tartu.
- Pajusalu, R. (1997). Eesti pronomeneid I. Ühiskeele *see*, *too*, *tema/ta* [Estonian pronouns I. *See*, *too*, *tema/ta* in common Estonian]. *Keel ja Kirjandus*, 1, 24–30; 2, 106–115.
- Palmeos, P. (1981). *Eesti keele grammatika II. Morfoloogia. Arv- ja asesõna*. [Estonian grammar II. Morphology. Numbers and Pronouns]. Tartu, Estonia, Tartu Riiklik Ülikool.
- Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. In J. H. Yoon & A. Kathol (Eds.), *OSU working papers in linguistics* (Vol. 49, pp. 91–136). Columbus, OH: Department of Linguistics, The Ohio State University.
- Romero, M., & Han, C.-h. (2003). Focus, ellipsis and the semantics of alternative questions. In C. Beyssade, et al. (Eds.), *Empirical issues in formal syntax and semantics 4* (pp. 291–307). Paris: Sorbonne.
- Rooth, M. (1985). *Association with focus*. Unpublished doctoral dissertation, University of Massachusetts, Amherst.

- Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, 1, 75–116.
- Saarimaa, E. A. (1949). Kielemme käytäntö. Pronominivirheistä [Language use. About pronoun errors]. *Virittäjä*, 49, 250–257.
- Sauerland, U. (2000). The content of pronouns: Evidence from focus. In T. Matthews & B. Jackson (Eds.), *Proceedings of SALT 10* (pp. 167–184). Ithaca, NY: CLC Publications/Cornell University.
- Seppänen, E.-V. (1998). *Läsnäolon pronominit. Tämä, tuo, se ja hän viittaamassa keskustelun osallistujaan* [Pronouns of presence. *Tämä, tuo, se* and *hän* referring to a conversational participant]. Helsinki, Finland: Suomalaisen Kirjallisuuden Seura.
- Stevenson, R., Knott, A., Oberlander, J., & McDonald, S. (2000). Interpreting pronouns and connectives. *Language and Cognitive Process*, 15, 225–262.
- Sulkala, H., & Karjalainen, M. (1992). *Finnish*. London: Routledge.
- Tael, K. (1988). *Sõnajärjemallid eesti keeles (võrrelduna soome keelega)* [Word order patterns in Estonian (in comparison with Finnish)]. Tallinn, Estonia: Eesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituut.
- Tauli, V. (1983). *Standard Estonian grammar. Part II syntax*. Uppsala: Acta Universitatis Upsaliensis, Studia Uralica et Altaica Upsaliensia.
- Umbach, C. (2004). On the notion of contrast in information structure and discourse structure. *Journal of Semantics*, 21, 155–175.
- Vallduví, E., & Vilkuna, M. (1998). On rheme and kontrast. In P. Culicover & L. McNally (Eds.), *The limits of syntax, syntax and semantics* (Vol. 29, pp. 79–108). New York: Academic.
- Varteva, A. (1998). Pronominit hän ja tämä tekstissä [The pronouns *hän* and *tämä* in text]. *Virittäjä*, 102, 202–223.
- Vilkuna, M. (1989). *Free word order in Finnish. Its syntax and discourse functions*. Helsinki: SKS.
- Vilppula, M. (1989). Havaintoja hän- ja he-pronominien käytöstä suomen murteissa [Observations about the use of the pronouns *hän* and *he* in Finnish dialects]. *Virittäjä*, 93, 389–399.
- Vsevirov, T. (1983). *Lausetevaheline pronominalisatsioon eesti keeles* [Cross-sentential pronominalization in Estonian]. Unpublished dissertation, Tartu, Estonia: University of Tartu.
- Zimmermann, M. (2007). Contrastive focus. In C. Féry, G. Fanselow, & M. Krifka (Eds.), *The notion of information structure* (pp. 147–159). Potsdam, Germany: Interdisciplinary Studies on Information Structure.