

Pronouns and Demonstratives in Finnish: Indicators of Referent Salience

Elsi Kaiser

University of Pennsylvania
Department of Linguistics
619 Williams Hall
36th and Spruce Street
Philadelphia, PA 19104-6305
ekaiser@babel.ling.upenn.edu

Abstract

This paper presents a corpus-based analysis of the referential properties of the pronoun *hän* ‘s/he’ and the anaphoric demonstrative *tämä* ‘this, s/he’ in Finnish, an SVO language with flexible word-order. Both forms can have human antecedents (Sulkala & Karjalainen, 1992). Following Gundel, Hedberg & Zacharski (1993), I assume that pronouns refer to entities at the “center of attention,” whereas demonstratives have less accessible referents. I analyze the antecedents of *hän* and *tämä* in a written corpus to determine whether the following factors correlate with referent salience: (a) Grammatical role: When the antecedent and anaphor are in separate main clauses, the antecedent of *hän* is usually the subject (71.67%, 43/60); *tämä*’s antecedent is often an object or adjunct (67.57%, 25/37). This supports the widespread view that subjects are more salient than objects. (b) Main-subordinate distinction: Referents in a subordinate clause are more often referred to with *tämä* than *hän* in the following matrix clause (*tämä*:77.27%, 17/22); *hän*:22.73%, 5/22) – suggesting that subordinate-clause referents are less salient than main-clause referents. (c) Word-order: According to native-speaker judgments, argument order does not correlate straightforwardly with salience. Different structures with identical argument order (Obj-Subj-verb, Obj-verb-Subj) have different pragmatic functions, which influence argument salience. This antecedent-distribution analysis indicates that, in Finnish, referent salience correlates with grammatical role and the matrix-subordinate distinction, but with not argument order.

1. Introduction*

In this paper I present a corpus-based analysis of the distribution and discourse functions of the pronoun *hän* ‘s/he’ (ex. 1) and the anaphoric demonstrative *tämä* ‘this’ (ex. 2) in Finnish, both of which can have human antecedents. (Boldface indicates coreference.)

(1a)

Sitten **eversti** piti puheen.... (p.144)
Then **colonel** held speech....
‘Then the **colonel** gave a speech.’.....

(1b)

...**Hän** koetti saada ääneensä tiettyä toverillista sävyä.
...**He** tried to-get voice-in-his certain friendly tone
....‘**He_{pro}**tried to get a certain friendly tone into his voice.’

(2)

Lammio huusi **Mielosta**, ja **tämä** tuli sisään lähetit kannoillaan. (p.286)
Lammio shouted-for **Mielonen**, and **this** came in messengers heels-on-his
‘Lammio called for **Mielonen**, and **he_{dem}** came in with the messengers on his heels.’

I claim that *hän* ‘s/he’ and *tämä* ‘this’ are used for referents with different levels of salience, and we can look at their distribution in order to find out more about what factors correlate with referent salience. On the basis of a corpus study and speaker intuitions, we can conclude that in Finnish, grammatical role and the difference between matrix and subordinate clauses correlate with salience, but the nature of the relation between word order and salience is not clear and merits further research.

This paper is structured as follows. In Section 1.1, we consider the notion of ‘referent salience’ and the factors which can influence it. Section 1.2 is a brief introduction to the main characteristics of the Finnish language, and in Section 2 I provide a brief overview of previous work on the Finnish anaphoric demonstrative. In Section 3, I provide data concerning the general distribution of pronouns and demonstratives in my corpus, and Section 4 is a discussion of the role that the grammatical role of the antecedent plays in anaphor choice. In Section 5 I address the distinction between main and subordinate clauses to see if it correlates with referent salience and anaphor choice, and in Section 6 I discuss the role of word order. Section 7 is the conclusion.

1.1 Referent salience

In this section we consider four factors which, according to existing work, are correlated with the salience of a referent: NP form, grammatical role, word order and the matrix-subordinate clause distinction. We will also explore the predictions these factors make concerning the pronoun-demonstrative distinction in Finnish. First, let us consider NP form. According to Gundel, Hedberg & Zacharski (1993), the form of the referring expression

* I would like to thank Miriam Eckert, Kimiko Nakanishi, Ellen Prince, Maribel Romero and Kieran Snyder for their valuable comments and suggestions on this paper. Earlier versions of this paper were presented at SECOL LXII (April 2000) and the University of Pennsylvania In-house linguistics colloquium (Spring 2000). Thanks to the audiences for comments and discussion. All errors are, of course, mine.

correlates with the cognitive status of the referent. Their analysis states that while pronouns are used to refer to entities that are “at the current center of attention” (Gundel, Hedberg & Zacharski, 1993:279), demonstratives are used for less accessible referents. I will follow Gundel, Hedberg & Zacharski (1993) and assume that in Finnish, the pronoun *hän* ‘s/he’ refers to entities which are more salient than those referred to by the demonstrative *tämä* ‘this.’

A link has also been claimed to exist between grammatical roles and salience. Brennan, Friedman & Pollard (1987), among many others, claim that subjects are more salient than objects. This, combined with the claims of Gundel, Hedberg & Zacharski (1993), predicts that *hän* ‘s/he’ will tend to refer to subjects, and *tämä* will tend to refer to objects. As we will see in Section 4, this is indeed the case for Finnish.

Another important question is whether word order is correlated with salience. The answer is not clear, and there is some evidence suggesting that the role played by word order may vary from language to language. On the one hand, Rambow (1993) argues that, in German, word order correlates with salience. He provides data illustrating that more salient entities occur to the left of less salient ones in the German middlefield. On the other hand, Turan (1998) and Hoffman (1998) claim that in Turkish, the salience of a referent correlates with its grammatical (or semantic) role, and is not affected by word order. They argue that in Turkish, subjects are more salient than objects even in scrambled sentences where the object linearly precedes the subject. These crosslinguistically different claims bring up the question; in Finnish, is the distribution of *hän* ‘s/he’ and *tämä* ‘this’ correlated with grammatical role or with word order? We will tackle this question in Section 6.

So far, we have discussed the connections between referent salience on the one hand and NP form, grammatical role and word order on the other hand. The fourth factor that we will consider is the Matrix-subordinate clause distinction. In other words, does the subject/object of a subordinate clause have a different level of salience than the subject/object of a main clause? According to Bever & Townsend (1979:167), main clauses are, in some sense, “more important” than subordinate clauses. They cite experimental evidence showing that adults and children process main clauses more “deeply” than subordinate clauses, and have “better access” to the meaning of a main clause (Bever & Townsend, 1979:176-177). In contrast, they remember the verbatim form of a subordinate clause better than that of the main clause (Bever & Townsend, 1979:176-177). This claim that main clauses are processed more deeply than subordinate clauses suggests that the entities mentioned in them are likely to be more accessible/salient than those mentioned in subordinate clauses. We will explore this prediction in Section 5.

1.2 Basics of Finnish

Finnish is a highly inflected language with flexible word order. It has canonical subject-verb-object order, but any of the six possible permutations of these three elements is grammatical in the appropriate context (Vilkuna 1995). Finnish has a gender-neutral pronoun *hän* ‘s/he,’ but the demonstrative *tämä* ‘this’ can also be used anaphorically in certain contexts, as illustrated in (1) and (2) above. In this paper, we will tackle two main questions: (a) in which contexts is *tämä* ‘this’ used, and in which contexts is *hän* ‘s/he’ chosen? and (b) what does this distribution pattern tell us about the factors influencing referent salience in Finnish?

2. Previous Work on the Finnish Demonstrative

In this section I provide a brief review of existing work on the anaphoric demonstrative *tämä* ‘this’ in Finnish, focusing both on the insights and shortcomings of previous research. Hakulinen & Karlsson (1988) suggest that the ‘demonstrative pronoun’ *tämä* refers to the most recently mentioned referent (Hakulinen & Karlsson, 1988:319). However, as illustrated by (3), this constraint alone is not sufficiently specific. In (3), where the most recently mentioned entity is also the only one mentioned so far, only the regular pronoun can be used to refer to Liisa.

(3)

Liisa nukkuu kotona. **Hän** / ??? **Tämä** on sairas.

Liisa sleeps home-at. **S/he** / ??? **This** is sick.

‘**Liisa** is sleeping at home. **She** is sick.’

Thus, a seemingly better way of capturing the use of *tämä* ‘this’ is to say, as Sulkala & Karjalainen (1992) do, that *tämä* ‘this’ is “used to indicate the last mentioned out of two or more possible referents” (282-283, emphasis added). This explains why *tämä* is not used in (3); only one possible referent is present.

However, this account leaves an important question unanswered: Is the demonstrative used to refer to last mentioned entity regardless of grammatical role? More specifically, what happens when the order of arguments is object-verb-subject (OVS) or object-subject-verb (OSV), instead of the canonical subject-verb-object (SVO)? An answer is offered by Saarimaa (1949), who argues that *tämä* ‘this’ refers to a recently mentioned, *non-subject* referent. Thus, it seems that, according to Saarimaa, the demonstrative cannot refer to the subject of a sentence even if the word order is OVS or OSV. However, in naturally-occurring language, is this really the case? We would also like to know whether subjects of matrix and subordinate clauses behave differently with respect to *tämä*.

In sum, the approaches discussed above leave a number of questions about the distribution of the pronoun and the demonstrative unanswered. Some answers are offered by a detailed, corpus-based analysis by Halmari (1994). She performed a large-scale corpus study of

referential expressions in written Finnish prose, and analyzed the distribution and antecedents of zero anaphors, pronouns, demonstratives, definite descriptions and proper names. She found that pronouns usually have subject antecedents (314/433, 72.5%), whereas the antecedents of demonstratives occupy a range of grammatical roles (see Table 1 below, from Halmari, 1994:53). However, although her corpus contained 433 pronoun tokens, it included only 15 demonstrative tokens. Halmari herself notes that “the huge number of pronouns in the sample skews the percentages, and this is a problem that needs to be addressed in future search...” (Halmari, 1994:55). In addition to the skewing influence of the large number of pronouns, it is not clear whether reliable results can be obtained from such a small number of demonstrative tokens.

	<i>hän</i> (pro)	<i>tämä</i> (dem)
subject	314 (72.5%)	2 (13%)
direct object	26 (26%)	4 (27%)
indirect object	7 (1.5%)	2 (13%)
oblique	31 (7%)	4 (27%)
genitive	55 (13%)	3 (20%)
object of comparison	-	-
Total	433 (100%)	15 (100%)

Table 1: Referring expressions and grammatical roles of immediate antecedent (recreated from Halmari 1994:53)

Putting aside for a moment the difficulties of the small token size, let us consider Halmari’s conclusions. She suggests that the distribution of *hän* ‘s/he’ and *tämä* ‘this’ indicates that there exists a correlation between the grammatical role of the antecedent and the choice of anaphoric NP. According to her, the pronoun *hän* refers to subjects, i.e. to highly salient entities, and the demonstrative *tämä* often refers to objects, which are less salient. This is not surprising, and it is what we would predict on the basis of Gundel, Hedberg & Zacharski (1993) and Brennan, Friedman & Pollard (1987), *inter alia*.

However, after a review of the existing literature, a number of questions still remain open. First, we would like to know whether using a larger corpus of demonstratives reveals clearer tendencies in the distribution of their antecedents, especially their grammatical roles. In Sections 3 and 4, I provide evidence illustrating that this is indeed the case. More specifically, my analysis of a corpus of written Finnish supports Halmari’s suggestion that the choice of anaphor form correlates with the grammatical role of the antecedent. Second, does the distinction between matrix and subordinate clauses influence the use of *hän* ‘s/he’ and *tämä* ‘this’? In Section 5 I show how entities in subordinate clauses are more likely to be referred to with demonstratives – i.e. are less salient – than entities in main clauses. Third, in Finnish, does word order variation influence the use of *hän* ‘s/he’ and *tämä* ‘this’? In response to this question, in Section 6 we will see that

it seems that word order interacts with a number of other factors when it comes to anaphor choice, but its relation to salience is not clear.

3. General Distribution of Pronouns and Anaphoric Demonstratives

In this section I discuss the general tendencies present in the distribution of *hän* ‘s/he’ and *tämä* ‘this’ in my corpus. My corpus consists of 103 occurrences of *hän* ‘s/he’ and 101 occurrences of *tämä* ‘this’ in the novel *Tuntematon Sotilas* ‘Unknown soldier’ by Linna (1954, 1999). I coded each anaphor for the following variables: its grammatical role, the word order of the sentence in which it occurs, the grammatical role of its antecedent, the word order of the sentence in which the antecedent occurs, as well as the matrix/subordinate¹ status of the sentence containing the antecedent and the sentence containing the anaphor.

Generally speaking, out of all the cases where an anaphoric element has its antecedent in a matrix clause, the pronoun *hän* is used to refer to the antecedent 60.26 % (91/151) of the time, and the demonstrative *tämä* is used 39.73 % (60/151) of the time. In contrast, out of all the cases where an anaphoric element has its antecedent in a subordinate clause, the demonstrative *tämä* ‘this’ is used 77.27 % (17/22) of the time, and the pronoun *hän* ‘s/he’ is used 22.72 % (5/22) of the time. Thus, even at this simplified level of analysis, it seems that matrix and subordinate clauses pattern differently. This will be discussed in more detail below.

Looking at the data from the perspective of the two different anaphors, I found that, overall, the pronoun *hän* tends to have its antecedent in the preceding matrix clause (88.35 %, 91/103). Its antecedent is very rarely in a subordinate clause (4.85 %, 5/103). In addition, the antecedents of the anaphoric demonstrative *tämä* also tend to be located in a preceding matrix clause (60/101, 59.41%), but in 16.83% (17/101) of the cases, they are in a subordinate clause. The general tendency of both anaphors to refer to antecedents in a matrix clause is due to the fact that, in my corpus, anaphoric reference is made more frequently to entities mentioned in a matrix clause than in a subordinate clause. In order to control for this bias and to see if there are more fundamental matrix-subordinate differences in the distribution of the anaphors or their antecedents, we need to look at the data in more depth. This is done in Section 4.

A more complete picture of the general distribution of the anaphors and their antecedents is given in Table 2. There are five main configuration types in which *hän* ‘s/he’ and *tämä* ‘this’ occur. First, we will consider the MM (matrix-matrix) configuration, in which

¹ In this paper, I use the term ‘subordinate clause’ as a general label to refer to all types of embedded clauses, including complement clauses (including Finnish nominalized subordinate clauses), relative clauses and adverbial clauses. I plan to address the differences between these various clause types in future research.

the antecedent and the anaphor are in two different main clauses. An example is “**Peter** went to the library yesterday. **He** needed to return some books.” This is the configuration in which both *hän* ‘s/he’ and *tämä* ‘this’ occur most frequently (60/103, 58.25% and 37/101, 36.63% respectively, shown in Table 2).

Another possible situation is the M(s).M / M.(s)M configuration. Here, the antecedent is in a main clause (M), and the anaphor is in a subsequent main clause (M). However, the two matrix clauses are separated by a subordinate clause (s, belonging to one or the other) that does not mention the antecedent. For example, “**Peter** thought that school would be cancelled today. **He** was wrong.” *Tämä* ‘this’ rarely occurs in this configuration (1/101, 1%), whereas it is the second most frequent configuration for *hän* ‘s/he’ (18/103, 17.48%).

In the third configuration, called Ms, the antecedent is in a main clause, and the anaphor is in a subordinate clause belonging to the main clause, e.g. “**Peter** said that **he** was tired.” Both *hän* ‘s/he’ and *tämä* ‘this’ occur in this configuration approximately 12 % of the time (12/103 and 12/101 respectively).

The fourth configuration type can be represented as (M)s.M. Here, the antecedent is in a subordinate clause belonging to a preceding main clause, and the anaphor is a subsequent main clause. For example, “Peter said [that **John** was sick yesterday]. **He** stayed in bed all day.” As Table 2 indicates, this configuration is significantly more frequent with *tämä* ‘this’ (17/101, 16.83%) than with *hän* ‘s/he’ (5/103, 4.85%).

The fifth and last configuration type, M.(M)s, has the antecedent in a main clause, and the anaphor in a subordinate clause belonging to the *next* main clause (which does not mention the antecedent). An example of this configuration is “**Peter** was really happy yesterday. The math teacher said [that **he** was really smart].” *Tämä* ‘this’ (10/101, 9.9%) occurs in this configuration much more frequently than *hän* ‘s/he’ (1/103, 1%).

	Hän (pro)	Tämä (dem)
a. MM	60 (58.25%)	37 (36.63%)
b. M.(s)M, M(s).M	18 (17.48%)	1 (1%)
c. Ms	12 (11.65%)	12 (11.88%)
d. (M)s.M	5 (4.85%)	17 (16.83%)
e. M.(M)s	1 (1%)	10 (9.9%)
Other	7 (6.79%)	16 (15.84%)
Not categorized	-	8 (7.93%)
Total	103	101

Table 2: General distribution

- None of the anaphors were ambiguous.
- Antecedent means immediate antecedent. M=main clause; s=subordinate clause.
- Parentheses indicate that that constituent does not contain the immediate antecedent.

So far we have seen that entities mentioned in a main clause can be referred to with either the pronoun *hän* or the demonstrative *tämä* (e.g. in the MM and the Ms

configurations, (a) and (c) in Table 2). This prompts the question: if an entity is mentioned in a main clause, are *hän* ‘s/he’ and *tämä* ‘this’ both equally viable ways of referring back to this entity? In the next section we will see that the answer is no: the grammatical role of the antecedent influences the pronoun/demonstrative choice in these contexts. However, entities mentioned in a subordinate clause are more likely to be referred to (in the next main clause) with a demonstrative than with a pronoun. This suggests that main and subordinate clauses differ in terms of the salience of their constituents.

4. Grammatical Role of the Antecedent

In this section we take a closer look at the influence that the grammatical role of the antecedent has on the choice of the anaphoric expression. As we will see, data from the corpus suggests that the pronoun *hän* ‘s/he’ tends to have antecedents that are subjects, whereas the demonstrative *tämä* ‘this’ often has object antecedents.

4.1 Distribution of antecedents of pronoun *hän*

The pronoun *hän* ‘s/he’ shows a distinct tendency to refer back to a preceding subject. When both the antecedent and the anaphor are in main clauses (MM), the antecedent tends to be the subject of the clause (43/60, 71.67%), as shown in Table 3a. As shown in Table 3b, the same holds when the antecedent is in a main clause and the anaphor in a subordinate clause belonging to it (Ms) (7/12, 58.33%). Moreover, the pronoun *hän* tends to be the subject of its own clause, regardless of whether it is in a main clause (43/60, 71.67%) or a subordinate clause (8/12, 66.67%).

Role of antecedent	Number of occurrences
S	43 (71.67%)
Poss ²	10 (16.67%)
DO	1 (1.67%)
IO	3 (5%)
Oblique	3 (5%)
PP	-
Total	60

Table 3a: Grammatical role of antecedent of *hän* (MM)

Role of antecedent	Number of occurrences
S	7 (58.33%)
Poss	4 (33.33%)
DO	-
IO	-
Oblique	-
other	1 (8.33%)
Total	12

Table 3b: Grammatical role of antecedent of *hän* (Ms)

² ‘Poss’ stands for possessive/genitive forms, e.g. **Peter’s** book.

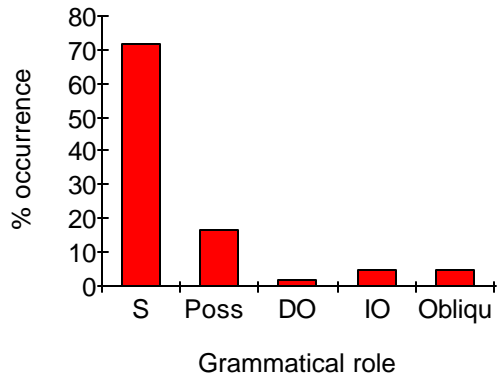


Chart 3a: Antecedents of pronouns (MM)

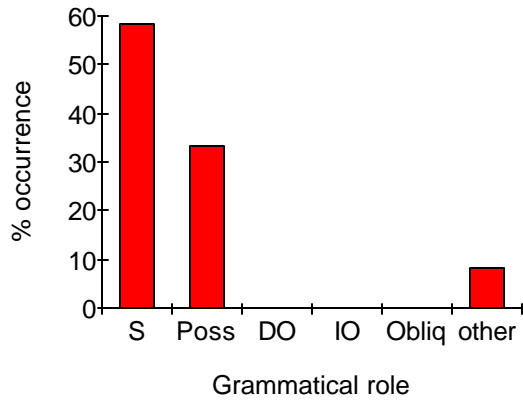


Chart 3b: Antecedents of pronouns (Ms)

4.2 Distribution of antecedents of anaphoric demonstrative *tämä*

Unlike the pronoun *hän* ‘s/he’, the demonstrative *tämä* ‘this’ usually has a non-subject antecedent. When both the antecedent and the anaphor are in main clauses, the antecedent tends to be the direct object of the clause (13/37, 35.14%), as shown in Table 4a. When the antecedent is in a main clause and the anaphor in a subordinate clause belonging to it, the same generalization obtains (6/12, 50%), as shown in Table 4b. In addition, the demonstrative *tämä* tends to be the subject of its clause, in both main (23/37, 62.16%) and subordinate clauses (5/12, 41.67%), but this is a weaker tendency than the one displayed by the pronoun *hän*.

Role of antecedent	Number of occurrences
S	7 (18.92%)
Poss	5 (13.51%)
DO	13 (35.14%)
IO	1 (2.70%)
Oblique	6 (16.22%)
PP	5 (13.51%)
Total	37

Table 4a: Grammatical role of antecedent of *tämä* (MM) c

Role of antecedent	Number of occurrences
S	-
Poss	1 (8.33%)
DO	6 (50%)
IO	1 (8.33%)
PP	3 (25%)
Other	1 (8.33%)
Total	12

Table 4b: Grammatical role of antecedent of *tämä* (Ms)

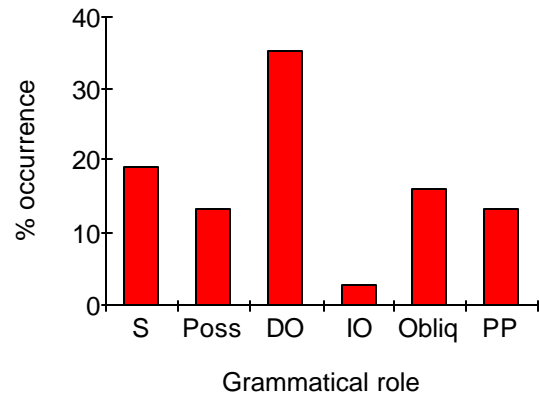


Chart 4a: Antecedents of demonstratives (MM)

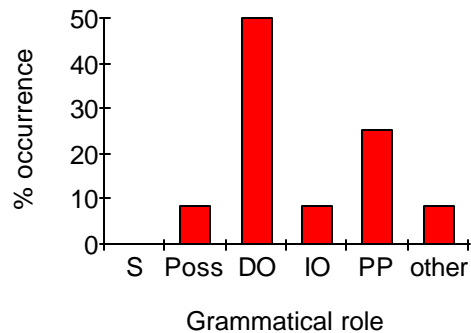


Chart 4b: Antecedents of demonstratives (Ms)

Thus, we can conclude that when both the antecedent and the anaphor are in main clauses (MM configuration), the pronoun often refers to the subject of the preceding main clause (Chart 3b), and the demonstrative tends to refer to an object of the main clause (Chart 4a). The same tendency exists in contexts where the antecedent is in a main clause and the anaphor is in a subordinate clause belonging to that main clause (Ms configuration). In more general terms, the data indicate that the choice of anaphor form – when referring to an antecedent in a main clause – correlates with the grammatical role of the antecedent.

5. Main and Subordinate Clauses

In the preceding sections, we have seen that, when the antecedent is in a matrix clause, the pronoun *hän* tends to refer to subjects and the demonstrative *tämä* to objects. In this section we will tackle the question, do these tendencies also hold for subordinate clauses? In this section we will focus primarily on the two configurations shown in (4) and (5):

(4) No potential referents in main clause (no ‘competitor’):

- (4a) The Johnsons noticed that **Peter** was sick.
(4b) **He** was coughing all night.

(5) Potential referents in main and subordinate clause (‘competing’ referents):

- (5a) Peter and John were sitting out on the balcony.
(5b) **Peter** declared that **John** should go to the Bahamas for a vacation.
(5c) **He** smiled happily.

In the configuration shown in (4), the anaphor is the subject of a separate sentence, but its antecedent is the embedded subject, and there is no potential antecedent (‘competitor’) in the main clause. The crucial question for configurations of this type is, what form of the anaphor (demonstrative or pronominal) is used to refer to the embedded subject? In (5), the anaphor is once again the subject of its own sentence, and its antecedent is the subject of the subordinate clause. However, this time a ‘competitor’ is present in the main clause. Thus, we want to find out what form of the anaphor is used to refer to the embedded subject, and what form is used to refer to the matrix subject. In both cases, we are also interested in the anaphor types used to refer to objects in the subordinate clause. In this section, I provide evidence showing that (a) when two or more potential antecedents are present (as in (5)), the demonstrative is used for the less salient referent, and (b) when only one possible referent is present (as in (4)), the demonstrative anaphor can still be used, even though a pronoun would be unambiguous.

5.1 Referring to arguments of subordinate clauses in the absence of competitors

First, let us consider the configuration shown in (4), i.e. which form is used to refer to the embedded subject in the absence of competitors? The corpus contains sentences in which this is done with a pronoun, and others in which the demonstrative is used. There are five cases where the matrix clause contains no ‘competing antecedent’ and the embedded subject is referred to with a pronoun (as in (6)).

(6a)

Asia oli nimittäin sillä tavoin, että **Lehto** oli saanut kapteenilta luottamustoimen. (p.13)

Thing was namely that way, that **Lehto** had received captain-from responsible-job.

‘The thing was that **Lehto** had been given, by the captain, a job demanding responsibility.’

(6b)

Hän oli huolehtinut kapteenin perheen muutosta kaupungissa. (p.13)

He had taken-care of captain’s family’s move city-in
‘**He**_{pro} had taken care of the captain’s family’s move in the city.’

In addition, the corpus contains two examples where the embedded subject is referred to with the demonstrative in the next main clause, and one example where an embedded object is referred to with the demonstrative. Interestingly, in all three cases, use of the pronoun would not create an ambiguity.

(7a)

Hietanen ja Vanhala huomasivat, että **Rokka** tarkoitti totta (p.231)

Hietanen and Vanhala noticed, that **Rokka** meant truth
‘Hietanen and Vanhala noticed that **Rokka** was serious’

(7b)

ja yhdessä he kaappasivat **tämän** jälkeen. (p.231)

and together they took-off **this**’s after

‘and together they took off after **him**_{dem}.’

Thus, in cases where the matrix clause contains no competitor, embedded subjects can be referred to with the pronoun *or* the demonstrative. What distinguishes these two options? One difference appears to be the presence vs. absence of an animate referent in the main clause. In the cases where a pronoun is used, there is no animate referent in the main clause. When the demonstrative is used, even though there is no competitor in the main clause, there is another animate referent present (plural human referent). Thus, it may be the case that in these kinds of contexts, the demonstrative is used to emphasize its antecedent’s ‘lower salience’ relative to another animate referent.

5.1.1 Other contexts where the demonstrative in the absence of a competitor

There are two other contexts as well where the demonstrative is sometimes used even though a pronoun would be unambiguous: (a) when the antecedent is a postverbal subject and (b) when it is embedded inside prepositional phrase. It seems that in these cases the demonstrative is used to refer to a referent that is less salient than another referent in the surrounding context. First, let us consider cases where the antecedent is embedded, such as (8).

(8a)

Käsikranaatti putosi muutaman metrin päähän

Koskelasta. (p.83)

Handgrenade fell few meter away-from **Koskela.**

‘The handgrenade landed a few meters away from **Koskela.**’

(8b)

Tämän päästä oli pudonnut lakki. (p.83)

this’s head-from had fallen hat.

‘The hat had fallen from **his**_{dem} head.’

Here, the antecedent ‘Koskela’ is located inside a prepositional phrase. In the next sentence, it is referred to with *tämä* ‘this’, even though *hän* ‘s/he’ would be perfectly unambiguous. Perhaps the demonstrative is used to illustrate the fact that, in the first sentence, the handgrenade is the most salient entity, and Koskela is much less accessible.

Another context in which demonstrative are used without a competitor present is when referring to postverbal subjects. Out of 11 postverbal subjects in my corpus (in the MM configuration), 7 were referred to with the pronoun *hän*, and 4 with the demonstrative *tämä*. No competing arguments were present in any of the examples. In example (9), the noun phrase *mies* ‘man’ is the subject of its sentence but occurs postverbally. In Finnish, postverbal subjects are usually indefinite (new information) (Chesterman, 1991), and thus one might conjecture that they are less salient than the old information present in the same sentence (such as the ‘corner of the trench’ in (9a)). If we accept this line of reasoning, then it is not surprising that in (9b), the demonstrative pronoun is used to refer to the man introduced postverbally in the previous sentence.

(9a)

Yhdyshaudan kulman takaa häämötti **mies**, (p.331)

Trench’s corner’s behind was-vaguely-visible **man**

‘Behind the corner of the trench **a man** was dimly visible’

(9b)

ja vain silmänräpäyksen **tämä** ehti epäröidä.... (p.331)

and only eyeblink **this** had-time to-hesitate...

‘and **he**_{dem} only had a moment to hesitate....’

In sum, these examples suggest that in some contexts, the demonstrative is used – instead of the pronoun – in order

to mark the antecedent as being less salient than something else present in the context.

5.2 Referring to arguments of subordinate clauses in the presence of competitors

Let us now return to the two configurations sketched above. We saw in the preceding section that when the antecedent is in a subordinate clause and no competitor is present, either the demonstrative or the pronoun can be used to refer to it. However, when a competitor is present in the main clause and an anaphor is used to refer to an argument in the subordinate clause, what happens? The corpus contains one example where, despite the presence of a competitor in the main clause (genitive of subject), a pronoun is used to refer to the subject of the subordinate clause. The more common choice in this type of situation, however, is the demonstrative. The corpus contains 14 examples where the demonstrative is used to refer to an argument in the subordinate clause when a competing antecedent is present in the main clause, as illustrated below for an embedded subject.

(10a)

Vääpeli katseli ajatuksissaan eteiseen, jossa **kirjuri** kampasi tukkaansa. (p.23)

Sergeant looked in-thought vestibule-to, where **scribe** combed hair-his

‘Deep in thought, the sergeant looked towards the vestibule, where the **scribe** was combing his hair.’

(10b)

Tämä ilmehti peilin edessä.... (p.23)

This made-faces mirror’s in-front

‘**He**_{dem} was making faces in front of the mirror.’

These data illustrate that the demonstrative can be used to refer to entities in a subordinate clause when the matrix clause contains a ‘competitor’ - which is what we would expect if *tämä* points to the less salient of two possible antecedents.

5.3 Distribution of subordinate-clause antecedents of the demonstrative *tämä*

In the previous section we saw data suggesting that subordinate-clause arguments are less salient than main clause arguments. This distinction between main and subordinate-clause arguments is supported by the general distribution of subordinate-clause antecedents of the demonstrative. When the antecedent of *tämä* is in a subordinate clause, and the anaphor itself is in a subsequent main clause, the antecedent is equally likely to be a subject, a genitive noun or a direct object (Table 5). Comparing this finding with the observations that when the antecedent is in a main clause, (i) the demonstrative tends to refer to objects more often than to subjects and (ii) subjects are usually referred to with a pronoun, shows that subordinate clauses pattern differently from main clauses.

Role of antecedent	Number of occurrences
S	5 (29.41%)
Poss	5 (29.41%)
DO	5 (29.41%)
IO	1 (5.88%)
PP	1 (5.88%)
Other	-
Total	17

Table 5: Grammatical role of antecedent of *tämä* (s.M)

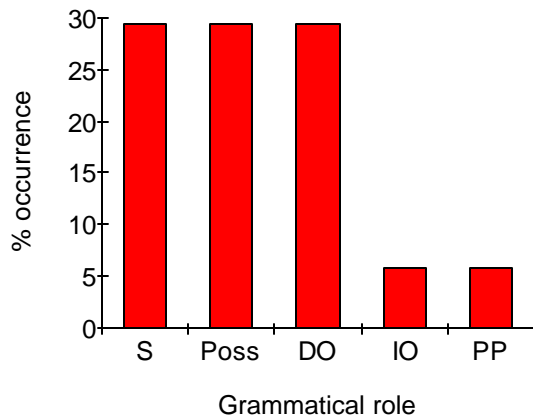


Chart 5: Antecedents of demonstratives (s.M)

6. Word Order

In this section we will address the question whether this distinction between subjects and objects is due to their grammatical role, or simply the fact that subjects occur before objects in the sentence. The crucial test case for this question is provided by sentences where the object precedes the subject (OVS, OSV). In the remainder of this section, we will explore the correlation between word order and anaphor choice.

6.1 Corpus data

Unfortunately there are no examples of the relevant type in the corpus, i.e. contexts where a third person singular animate object precedes a third person singular animate subject (OSV, OVS) and a pronoun or demonstrative is used to refer to one of them in the following clause.

6.2 Native speaker judgements

Another way of investigating the role of word order is via judgments. Halmari (1994) presented seven native speakers with OVS sentences out of context, and tested whether a pronoun/a demonstrative as the subject of a following sentence referred to the object or the subject of the OVS sentence (ex. (11)).

(11) OVS order

Kanan näki kissa ja **tämä** kuoli.

Chicken-object saw cat-subject and **this** died.

‘The cat saw the chicken and **it_{dem}** died.’ (Halmari 1994:42)

Halmari reports that native speakers had great trouble processing this sentence, but did show a very slight preference to interpret the demonstrative *tämä* as referring to the last mentioned entity, i.e. the subject (Halmari, 1994:42).

A possible reason for the processing difficulties experienced by the informants could be the isolated nature of the sentences. They were presented without any context, although OVS order is, in reality, only felicitous in particular pragmatic contexts (see Chesterman, 1991). Thus, it might be the case that the processing difficulties are a result of the pragmatic vacuum in which the sentences occur. It would be useful to see how people perform at this type of task when the sentences are given in a felicitous context.

In addition to the lack of context, the informants’ judgments of the sentences may be biased due to recency effects. Arnold (2000) found that in an English-language story continuation task, having an object at the very end of a sentence (without any lexical material after it) biases people to refer to it in their continuation. When the object is followed by a locative PP, a time phrase etc., people are less likely to continue with it. Crucially, in naturally-occurring language (corpus study), continuations are more likely to refer to subjects than objects. This is taken as an indication that subjects are more salient than objects.

Thus, Arnold’s findings suggest that certain experimental tasks, when not appropriately designed, may induce recency effects that do not match the tendencies observed in naturally-occurring language. This may be happening with sentences such as ex. (11); here, the subject is at the very end of the first phrase, which might be biasing people to refer to it when they encounter an anaphor.

When dealing with the relative order of subject and object, one also needs to remember that, in addition to OVS, which Halmari tested, in Finnish OSV is another word order possibility where the object precede the subject. However, OVS and OSV differ in their pragmatic interpretations: whereas OVS can be used when referring to an old object and introducing a new subject, OSV order is often associated with a contrastive interpretation (see Vilkuna, 1995), as illustrated in (12).

(12) OSV order

Opettajan tyttö tapasi (eikä poikaa).

Teacher-object girl-subject met (not boy-object)

‘The girl met a/the teacher (, not a/the boy).’

I found that native speaker judgements for pronouns and demonstratives referring to contrastive OSV sentences (with context) tend to vary: Preliminary findings suggest that a pronoun can refer to the preposed, contrastive

object, or to the subject, depending on the speaker. The demonstrative, however, is not interpreted as referring to the subject. In any case, there is no reason to expect OVS and OSV order to pattern in the same way when it comes to anaphoric reference, since their pragmatic properties are so different. This suggests that a statement such as 'tämä' refers to the second-mentioned referent' will probably turn out to be an oversimplification. Further research is clearly needed in this area.

In sum, the role of word order in the use of *hän* 's/he' and *tämä* 'this' is unclear. It seems that relying purely on linear order may be too simplistic an approach. The pragmatic functions of different word orders, as well as the context of the utterance, etc. may influence the salience of the arguments and hence the choice of the anaphor.

7. Conclusions

In this paper I used a corpus of written Finnish to explore the distribution of the pronoun *hän* 's/he' and the demonstrative *tämä* 'this.' My findings suggest that when referring to an entity in a main clause, the pronoun tends to refer to the subject (highly salient) and the demonstrative to the object (less salient). This provides support for Brennan, Friedman, Pollard (1987)'s claim that subjects are, in general, more salient than objects.

In addition, I claim that the correlation between word order and salience is not yet clear, and may depend on the particular discourse function(s) of the different word orders. The data also shows that, even though the demonstrative is not usually used to refer to matrix subjects, it can refer to an embedded subject – especially when a 'competitor' antecedent is present in the main clause. This suggests that main subjects are more salient than embedded subjects. The observation that the demonstrative is used in certain contexts in which a pronoun would also be unambiguous suggests that perhaps *tämä* can be used to mark its antecedent as being less salient than some other entity in the surrounding context.

As mentioned above, the role of word order is not yet clear, and further research is needed in that area. In addition, in this paper, all types of subordinate clauses are treated in the same way. However, it seems very possible that different types of subordinate clauses pattern differently and have differing levels of salience. Thus, to fully understand the matrix-subordinate distinction, it will be necessary to explore different kinds of subordinate clauses in detail.

References:

Arnold, J. (2000). The Effect of Thematic Roles on Pronoun Use and Frequency of Reference Continuation. *Penn Working Papers in Linguistics*, 6.3, Current Work in Linguistics, ed. A. Williams *et al.*
 Bever, T. & Townsend, D. (1979). Perceptual Mechanisms and Formal Properties of Main and Subordinate Clauses. In W. Cooper & E. Walker (eds).

Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett. Hillsdale, NJ: Lawrence Erlbaum.
 Brennan, S.E., Walker, M.A. & Pollard, C.J. (1987). A Centering Approach to Pronouns. *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, Stanford, Calif., 155-162.
 Chesterman, A. (1991). *On definiteness*. Cambridge, England: Cambridge University Press.
 Gundel, J.K., Hedberg, N. & Zacharski, R. (1993). Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69, 274-307.
 Hakulinen, A. & Karlsson, F. (1988). *Nykysuomen Lauseoppia*. Helsinki: SKS.
 Halmari, H. (1994). On Accessibility and Coreference. *Nordic Journal of Linguistics*, 17, 35-59.
 Hoffman, B. (1998). Word Order, Information Structure and Centering in Turkish. In M.A. Walker, A.K. Joshi, E.F. Prince (eds.), *Centering Theory in Discourse*. Oxford: Clarendon Press. 251-272.
 Linna, V. (1954/1999). *Tuntematon Sotilas*. Helsinki: WSOY
 Rambow, O. (1993). 'Pragmatic Aspects of Scrambling and Topicalization in German.' Paper presented at the Workshop on Naturally-Occurring Discourse, Institute for Research in Cognitive Science, University of Pennsylvania.
 Saarimaa, E.A. (1949). Kielemme käytäntö. Pronominivirheistä. *Virittäjä* 49. 250-257.
 Sulkala, H. & Karjalainen, M. (1992). *Finnish*. New York: Routledge.
 Turan, U. (1998). Ranking Forward-Looking Centers in Turkish: Universal and Language-Specific Properties. In M.A. Walker, A.K. Joshi, E.F. Prince (eds.), *Centering Theory in Discourse*. Oxford: Clarendon Press. 139-160.
 Vilkuna, M. (1995). Discourse Configurationality in Finnish. In Kiss, K. (ed.) *Discourse Configurational Languages*. New York: Oxford University Press. 244-268.
 Walker, M., Joshi, A. & Prince, E.F. (1998). Centering in Naturally Occurring Discourse: An Overview. In M.A. Walker, A.K. Joshi, E.F. Prince (eds.), *Centering Theory in Discourse*. Oxford: Clarendon Press. 139-160.