Word order, grammatical function, and referential form: On the patterns of anaphoric reference in Finnish[*]

Elsi Kaiser
University of Pennsylvania

1. Introduction

Previous research has shown that there exists a connection between the form of a referring expression and the accessibility/salience of its referent. More specifically, the most salient referents – i.e. those referents that are currently at the center of attention and most prominent at that point in the discourse – are referred to with the most reduced referring expressions (e.g. pronouns in English, null *pro* in Spanish). This raises the question: What kinds of factors influence a referent's salience, i.e. make it a good candidate to be referred to with a reduced anaphoric expression?

In this paper, I focus on two factors which have been claimed to influence referent salience: (1) grammatical/syntactic role and (2) word order. I will address these issues from the perspective of Finnish, a highly inflected, flexible word order language with canonical SVO order (Vilkuna 1995) and two kinds of third person anaphors: the gender-neutral pronoun *hän* 's/he' and the demonstrative *tämä* 'this.' I present the results of three psycholinguistic experiments investigating the referential properties of these two anaphors, and show that *hän* and *tämä* differ in their referential properties and are sensitive to different kinds of factors. The results indicate that instead of trying to define the referential properties of these forms according to a unified notion of salience, we should investigate how different factors may be relevant for different referential expressions.

The structure of this paper is as follows. First, in section 2, I review existing work on referent salience. In section 3, I discuss the word order patterns and anaphoric system of Finnish. Section 4 presents the results of the sentence-completion experiments and discusses their implications. Conclusions and directions for future work are addressed in section 5.

2. Salience and the form of referring expressions

There exists a general consensus that the more reduced an anaphoric expression is, the more salient its antecedent has to be, and that "pronouns are used most often when the referent is represented in a prominent way in the minds of the discourse participants, but more fully specified forms are

---

needed when the representation of the referent is less prominent" (Arnold 1998:4). This correlation is encoded in various accessibility hierarchies of referential forms that have been proposed in the literature (see e.g. Gundel, Hedberg and Zacharski 1993, Givón 1983 and Ariel 1990). According to these hierarchies, overt pronouns are used for more accessible antecedents than demonstratives, and null pronouns for more accessible referents than overt pronouns, and so on. Let us now take a look at two of the factors that have been claimed to have an impact on referent accessibility; word order and syntactic role.

2.1 Syntactic function

Previous research has found a close connection between grammatical roles and salience – specifically, that subjects are more salient than objects (e.g. Brennan, Friedman & Pollard 1987, Matthews & Chodorow 1988, Stevenson et al. 1994). Many researchers use anaphoric elements as a window to gain insight into what makes referents salient, and thus rely on the finding that the most reduced anaphoric element in a given language refers to the most salient referent. For example, Crawley & Stevenson (1990) conducted a sentence continuation experiment where participants were asked to continue stories like "Shaun led Ben along the path and he….". The continuations were analyzed to see how people interpret the pronoun *he*, which is assumed to refer to the most salient entity. The results indicate that the pronoun is interpreted as referring back to the subject significantly more often than to the object. The same subject advantage was found in reading-time studies (e.g. Gordon, Grosz & Gilliom 1993) and corpus studies (e.g. Brennan, Friedman & Pollard 1987, Tetreault 2001).

In sum, referents in subject position seem to be 'default topics' in that they are more likely to be the antecedents of pronouns than entities in other grammatical positions. However, with languages like English that have relatively rigid subject-object order, it is difficult to tell whether the increased salience of subjects is due to their position at the beginning of the sentence, or their semantic/thematic properties.

2.2 Word order

To untangle the notions of subjecthood and first position, and to see whether word order itself influences salience, we can turn to languages with flexible word order. Existing research reveals different findings for different languages. For example, for German, Rambow (1993) and Strube & Hahn (1996) found that word order correlates with salience (see also Lenerz 1977, Choi 1996 on German word order) and guides pronoun resolution, whereas 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. Let us first consider Rambow's German examples in (1a,b). According to Rambow, the pronoun in the answer tends to refer to the leftmost constituent in the *Mittelfeld*. Thus, when the constituent order is changed from subject-object (1a) to object-subject (1b), the preferred referent of the pronoun in (1c) changes.

(1)   a. subject-object order
Glauben Sie, dass [eine solche Maßnahme]$_a$ [der russischen Wirtschaft]$_b$
Think   you that [a such measure]-NOM   [the Russian economy]-DAT
helfen kann?
help   can?
'Do you think that such a measure can help the Russian economy?'

b. object-subject order
Glauben Sie, dass [der russischen Wirtschaft]$_b$ [eine solche Maßnahme]$_a$ helfen kann?

c. *answer*:
Nein, **sie** ist viel   zu   primitiv.
No,   she is   much too primitive.
'No, it (she) is much too primitive.'
(1a) it=measure (1b) it=economy

In contrast, in the Turkish counterparts to Rambow's examples, discussed by Turan (1998:142), the null pronoun in the answer is interpreted as referring to the subject, regardless of whether the word order is subject-object or object-subject. In other words, the claim is that in Turkish, subjects are more salient than objects even in scrambled sentences where the object linearly precedes the subject.

When faced with this kind of seemingly conflicting data, it is worth keeping in mind that the functions of scrambling vary across languages, and even in different constructions within a single language. In my opinion, it is reasonable to hypothesize that this may be at least part of the reason for the crosslinguistically conflicting findings. In fact, Rambow (1993) shows that, in German, 'topicalized' word orders sometimes have an impact on salience and at other times they do not. He argues that whether salience is determined by word order depends on the discourse function of the

topicalization. Thus, before concluding that languages differ in terms of whether word order variation influences salience, we need to consider the functions of different word orders in those languages. In the next section, we will turn to the discourse functions of SVO and OVS order in Finnish.

3. Finnish

The grammatical properties of Finnish make it a good testing ground for investigating how word order and grammatical role influence a referent's chances of being referring to in subsequent discourse with a reduced anaphoric expression. Finnish has flexible word order, and two kinds of third person anaphors (pronoun *hän* 's/he' and demonstrative *tämä* 'this, s/he'). We will investigate how word order and grammatical role affect the referential properties of these anaphors.

3.1 Finnish word order

Finnish has free word order and no definite or indefinite article.[1] The canonical word order is SVO, but all six permutations of these elements are grammatical in the appropriate contexts (Vilkuna 1995:245). In this paper, we will focus on SVO and OVS orders.

In order to understand why one might expect the SVO/OVS variation in Finnish to have an impact on referent salience, it is necessary to understand the pragmatic factors guiding the alternation. In Finnish, the choice between SVO and OVS order is guided by the discourse-status of the arguments, i.e. whether or not they have been mentioned in the preceding discourse. If we combine this observation with a claim by Strube & Hahn (1996) that discourse-status determines salience (i.e. discourse-old entities are more salient than discourse-new ones), we predict that the Finnish SVO/OVS variation determines the subject and object's salience.

Before we turn to the experiments to see if this prediction is supported, let us first consider the discourse properties of subjects and objects in SVO and OVS orders in more detail. First, let's look at subjects. Noncanonical, postverbal subjects introduce referents that are discourse-new, i.e. have not yet been mentioned in the current discourse.[2] This is illustrated by example

---

[1] In dialects of spoken Finnish, the demonstrative pronoun *se* 'it' is evolving into a kind of definite article (see Laury 1997). However, this does not occur in standard Finnish.

[2] In Finnish, the distinction between old and new information depends on the discourse status of the entities, not on whether or not they are known/old to the hearer (hearer-status). This is shown by the fact that names of family members or famous people (hearer-old) can surface as post-verbal subjects in Finnish, if they are discourse-new. (see ex. (4) below). See Prince (1992) for further discussion of the notions of discourse-status and hearer-status.

(2), where the English original has a sentence-initial indefinite noun phrase. In Finnish, there is no indefinite article, and the noun phrase occurs postverbally. In contrast, preverbal subjects are usually discourse-old information, i.e. refer to entities that have already been mentioned in the current discourse. This is exemplified by the Finnish translation in (3). A preverbal subject NP is interpreted as being new information only when the sentence is a discourse-initial 'all new' utterance.

(2)   postverbal subject
      a. **A great big water-beetle** came up underneath the lily leaf.
      b. Lumpeenlehden alla   ui   **iso**   **vesikuorianen**.
        *lily-leaf-GEN   under swam **big-NOM water-beetle-NOM**.*
    (from Beatrix Potter's *The tale of Mr Jeremy Fisher* (1979), Chesterman 1991:100).

(3)   preverbal subject
      a. **The tyrannosaur** was very close now. (Crichton, 240)[3]
      b. **Tyrannosaurus**   oli jo hyvin lähellä. (Finnish translation, 276)
        Tyrannosaur-NOM was already very close.

Let us now consider the discourse properties of objects. A preverbal object, followed by a postverbal subject, as illustrated in (4), is interpreted as discourse-old information. Postverbal objects can be interpreted as new *or* old information, as shown in (5).

(4)   OVS order
    Tiedotteen   välitti   julkisuuteen kurdien uutistoimisto DEM
    Announcement-ACC transmitted public-to Kurds' newsoffice-NOM DEM
    '**The announcement** was made public by the Kurdish newsoffice D.E.M.'

                      (from the newspaper *Aamulehti* 3/16/1999)

(5)   SVO order
    Mies   huomasi **kissan**.
    Man-NOM noticed   cat-ACC
    'The man noticed **a/the cat**.'

---

[3] Crichton, Michael. 1995. *The Lost World.* New York: Knoepf. Finnish version: Crichton, Michael. 1996. *Kadonnut Maailma.* Helsinki: Otava. (Finnish translation by Jaakko Kankaanpää)

3.2 Finnish anaphoric paradigm
In the previous section, we reviewed the pragmatic characteristics of SVO and OVS order, and saw that these two orders differ in terms of the discourse status of the subject and the object. Now, keeping in mind Strube & Hahn's (1996) claim that discourse status determines salience, we will take a closer look at previous work on the referential properties of the two anaphoric forms *hän* 's/he' and *tämä* 'this.'

Previous work on the referential properties of *hän* 's/he' supports the crosslinguistic generalization that overt pronouns (in languages without null pronouns) refer to the most salient entities. The pronoun *hän* has been described as referring to the most central or 'foregrounded' character (Kalliokoski 1991) or to the character who is most important in a given situation or context (Vilppula 1989, *inter alia*). According to Saarimaa (1949), *hän* tends to refer to the subject of the preceding sentence because the subject is more in the 'foreground' than other referents.

The demonstrative *tämä* 'this' differs from the pronoun *hän* in that it can function as a proximal demonstrative and a deictic, in addition to being used to refer to human antecedents. The referential properties of *tämä* also differ from those of *hän*: whereas the pronoun is used for foregrounded characters, *tämä* has been described as referring to characters in the background (Varteva 1998). In more structural terms, Sulkala & Karjalainen (1992) note that *tämä* is "used to indicate the last mentioned out of two or more possible referents" (1992:282-283). This raises the question: Does the demonstrative refer to the last mentioned entity regardless of grammatical role? What happens with OVS order? Saarimaa (1949) claims that *tämä* 'this' refers to a recently mentioned, non-subject referent and that *hän* is used for subject antecedents. However, the question remains: In actual language use and comprehension, is this the case?

A partial answer is offered by a corpus study of Finnish referential expressions by Halmari (1994). Her corpus contained 433 pronoun tokens, and 15 demonstrative tokens.[4] As she notes, "the huge number of pronouns in the sample skews the percentages, and this is a problem that needs to be addressed in future research" (Halmari 1994:55). As Table 1 (from Halmari 1994:53) shows, she found that *hän* refers to subjects, i.e. to highly salient entities, and the demonstrative *tämä* tends to refer to objects, which are less salient. Her corpus findings confirm the intuitions of other Finnish researchers – but do not give a conclusive answer to the question of how

---

[4] Halmari (1994) focused on a wide range of different referential expressions, and thus the pronoun-demonstrative distinction was not the primary focus of her investigation.

how word order affects the referential properties of *hän* and *tämä*, as she did not analyze word order in her corpus study.[5]

Table 1: Referring expressions and grammatical role of antecedent

	hän (pro)	*tämä* (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%)
Total	**433 (100%)**	**15 (100%)**

To address the imbalance of pronoun and demonstrative tokens in Halmari's corpus, I conducted a corpus study (Kaiser 2000) of 103 occurrences of *hän* 's/he' and 101 occurrences of *tämä* 'this' in the novel *Tuntematon Sotilas* 'Unknown soldier' by Linna (1954/1999, Helsinki:WSOY).[6] The results for *hän* are in Table 2. In general, *hän* 's/he' tends to refer to a preceding subject (43 out of 60 cases, 71.67%). In contrast, *tämä* tends to have a non-subject antecedent (Table 3). Examples are in (6) and (7) (bolded constituents are coreferential).

---

[5] Importantly, however, Halmari (1994) conducted a small survey and asked seven native speakers about sentences with different word orders and different anaphoric elements. She tested the OVS sentence *Kanan näki kissa ja {se/tämä} kuoli.* 'Chicken-ACC saw cat-NOM and {it/this} died.' People were given the sentence either with *se* 'it' or *tämä* 'this' and were asked 'Who died?' With the pronoun *se* 'it', there was a preference to interpret it as referring to the object *chicken* (presumably for pragmatic reasons, as a cat seeing a chicken is likely to result in the chicken dying, rather than the cat), and with the demonstrative *tämä* 'this', people did not give very clear responses and found the resulting sentence "extremely hard to process" (Halmari 1994:42).

[6] These data are for cases where the anaphor and its antecedent are in distinct main clauses. Subordinate clauses were also analyzed, but are not included here, so the totals shown here are less than 101 and 103. See Kaiser (2000) for details.

Table 2: Antecedent of *hän*   Table 3: Antecedent of *tämä*

Role of antecedent	Number of occurrences	Role of antecedent	Number of occurrences
S	43 (71.67%)	S	7 (18.92%)[7]
Poss[8]	10 (16.67%)	Poss	5 (13.51%)
DO	1 (1.67%)	DO	13 (35.14%)
IO	3 (5%)	IO	1 (2.70%)
Oblique	3 (5%)	Oblique	6 (16.22%)
PP	-	PP	5 (13.51%)
Total	60	**Total**	37

(6)   Example with *hän* (Linna:144)
Sitten **eversti**   piti puheen.   **Hän**   koetti saada
*Then **colonel-NOM** held speech-ACC. **He-NOM** tried   get*
ääneensä   tiettyä   toverillista   sävyä.
voice-into-his certain-PART friendly-PART tone-PART
'Then the **colonel** gave a speech. **He** tried to get a certain friendly tone into his voice.'

(7)   Example with *tämä* (Linna:286)
Lammio   huusi   **Mielosta**,   ja **tämä**   tuli sisään
*Lammio-NOM shouted **Mielonen-PART**, and **this-NOM** came in*
lähetit   kannoillaan.
messengers heels-on-his
'Lammio called for **Mielonen**, and **he** came in with the messengers on his heels.'

In sum, the results of both Halmari (1994) and Kaiser (2000) show that there is a correlation between anaphoric form and grammatical role. Subjects are usually referred to with *hän,* and objects and oblique arguments with *tämä*. Should we conclude, then, that subjects have a higher level of salience than objects or obliques? Not necessarily. We know

---

[7] The demonstrative *tämä* is used three times to refer to postverbal subjects, e.g.:

(i)   Yhdyshaudan kulman   takaa häämötti   **mies**,
*trench-GEN corner-GEN   behind was-vaguely-visible **man-NOM**,*
ja vain silmänräpäyksen **tämä**   ehti   epäröidä
*and only eyeblink-ACC   **this-NOM** had-time to hesitate*
'Behind the corner of the trench, a **man** was dimly visible, and **he** only had a moment to hesitate…' (Linna:331)

[8] 'Poss' stands for possessive/genitive forms, e.g. [**his** book], [**Peter's** book].

that in SVO order, the subject tends to be referred to with the 'salient anaphor' *hän*, but we don't yet know if this is due to linear order or grammatical function. To unconfound these factors, we need to look at the referential properties of *hän* and *tämä* for sentences where the object precedes the subject. However, finding sufficient numbers of such examples in an unparsed corpus is difficult, and the corpus I used for the Kaiser (2000) study did not a contain any examples of transitive verbs in OVS sentences with a human third person subject and object, followed by *hän/tämä*. To circumvent this problem, I used sentence completion tasks. In such experiments, participants are given sentences or sentence fragments and asked to provide natural-sounding continuations. The continuations are analyzed to see how participants interpreted the sentence provided. Below, I present three written sentence completion studies investigating how the grammatical function and linear position of potential antecedents influence the referential properties of *hän* and *tämä*.

4. Experiments

4.1 Experiment 1

This experiment tested the effect of word order and grammatical role on referential properties of *hän* and *tämä*. The stimuli consisted of written SVO and OVS sentences, each of which was followed by the first word of the next sentence, either *hän* 's/he' or *tämä* 'this.' Anaphor type and word order were crossed to create four conditions: [SVO.Hän...], [OVS.Hän....], [SVO.Tämä....] and [OVS.Tämä.....]. A participant's task was to write a completion for the second sentence. An example item is provided below.

(8)   [SVO.Hän]
     Lääkäri   onnitteli   opiskelijaa.   Hän...
     Doctor-NOM congratulated student-PART. S/he-NOM.....
     'A/the doctor congratulated a/the student. S/he...'

Sixteen native Finnish-speakers participated in this experiment. Each participant was asked to complete 38 items: 8 critical items and 30 fillers whose order was randomized. The nouns used for the subject and object in the critical items were all professions or other 'roles' (e.g. doctor, stewardess, reporter, student). This was done in order to make the continuations easier to interpret. All verbs used were action/agent-patient verbs (as defined by Stevenson et al. 1994). A unified verb group was used in order to control for any possible verb focusing effects. Continuations were coded according to which of the referents in the preceding sentence the participants chose as the referent of the pronoun.

The results are shown in Figure 1. As the graph shows, the referential properties of *hän* and *tämä* are affected in different ways by word order. There are significant effects of anaphor type and word order on reference to both subjects and objects. Analyses of variance show that whether an anaphoric element is interpreted as referring to the preceding subject depends on anaphor type (*hän* or *tämä*, $F1(1,15)=39.28$, $p<.01$) and word order (SVO or OVS, $F1(1,15)=7.75$, $p<.05$). There is also a significant interaction ($F(1,15)=6.48$, $p<.05$). Similarly, for objects, whether an anaphoric element refers to the preceding object is dependent on anaphor type ($F1(1,15)=12.42$, $p<.01$) and word order ($F1(1,15)=24.77$, $p<.01$). Again, there is a significant interaction ($F(1,15)=11.81$, $p<.01$).

Now, let us look in more detail at the four conditions. The pronoun *hän* 's/he' tends to be interpreted as referring to the subject, regardless of word order. Thus, in the [SVO.Hän] condition, the pronoun was interpreted as referring to the preceding subject in 23 out of 32 cases (72%). In the [OVS.Hän] condition, we also see 23/32 (72%) subject-interpretations. In contrast, in the [SVO.Tämä] condition, *tämä* tends to refer to the object in SVO order; it was interpreted as referring to the object in 26 out of 32 cases (81%). In the [OVS.Tämä] condition, order, however, *tämä* is split between the subject and the object. There are 14/32 (44%) subject-interpretations, and 9/32 (28%) object-interpretations. In this condition, the difference between the rate of subject-interpretations and object-interpretations is not statistically significant ($t(15)=.735$, $p>0.05$).

Figure 1. Which referent does the anaphor refer to?
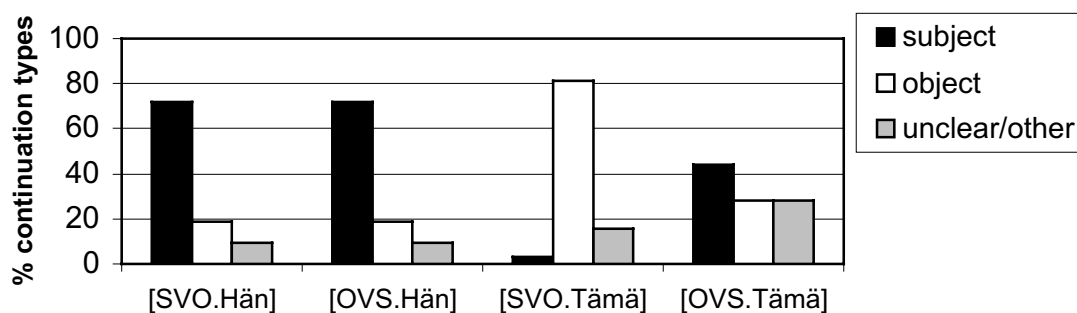
Before moving onto a discussion of these results, let's consider a related sentence-completion experiment that situated the sentences in discourse contexts (see Kaiser (in preparation)). Recall that the SVO/OVS variation in Finnish is driven by the discourse status of the arguments. Thus, in Experiment 1, the OVS sentences were infelicitous, because they were presented without a preceding context. In this second experiment,

Experiment 1b, a brief discourse context preceded the SVO/OVS sentences, such that the preverbal noun (S or O) was discourse-old, and the postverbal noun (O or S) was discourse-new.[9] The number of critical items was increased to 16, and they had the same structure as in Experiment 1: The nouns used were roles/professions, and only agent/patient verbs were used. Sixteen native Finnish speakers participated in this study.

The results of Experiment 1b exhibit the same patterns as we saw in the results of Experiment 1. The referential properties of *hän* and *tämä* are affected differently by word order. Importantly, the tendency seen in the first experiment in the [OVS.Tämä] condition for the demonstrative to prefer the postverbal subject over the preverbal object is statistically significant here (t(15)=3.91, p<.01). We can conclude that *tämä* prefers discourse-new postverbal subjects over discourse-old preverbal objects (see Kaiser (in preparation) for details).

Thus, these results show that the two referential forms are sensitive to different factors. The pronoun *hän* is sensitive to the syntactic function/grammatical role of potential antecedents and prefers subjects (see also Saarimaa 1949). In contrast, the demonstrative *tämä* is primarily sensitive to word order (which is correlated here with discourse status[10]). It prefers to refer to postverbal constituents, especially postverbal nonsubjects. In fact, I'd like to suggest that *tämä* is sensitive to salience – and since salience depends on factors such as word order/discourse status (e.g. Strube & Hahn 1996) and grammatical role (e.g. Crawley & Stevenson 1990), *tämä* is sensitive to these factors. More specifically, according to this hypothesis, *tämä* prefers entities that are low in salience, entities that are not at the center of attention at that point in the discourse (see also Varteva 1998). So, these results suggest that *tämä* is associated with the low-end of a salience scale, and *hän* with the high-end of a grammatical role scale. However, as we will see below, grammatical role is not the only thing that matters for *hän*.

---

[9] The contexts were created such that two full NPs could be felicitously used in the critical SVO/OVS sentence. This was done by means of two context sentences which mention a third referent (see Kaiser (in preparation) for details).

[10] *Tämä* can also refer to discourse-old referents, as is shown by corpus data. If it is preceded by a transitive sentence that contains two discourse-old arguments, which in Finnish will normally occur in S-O order, it prefers the object. However, if *tämä* is preceded with a sentence with one discourse-old and one discourse-new argument, which will tend to occur in old-before-new order, then, as the present results show, it will prefer the discourse-new referent.
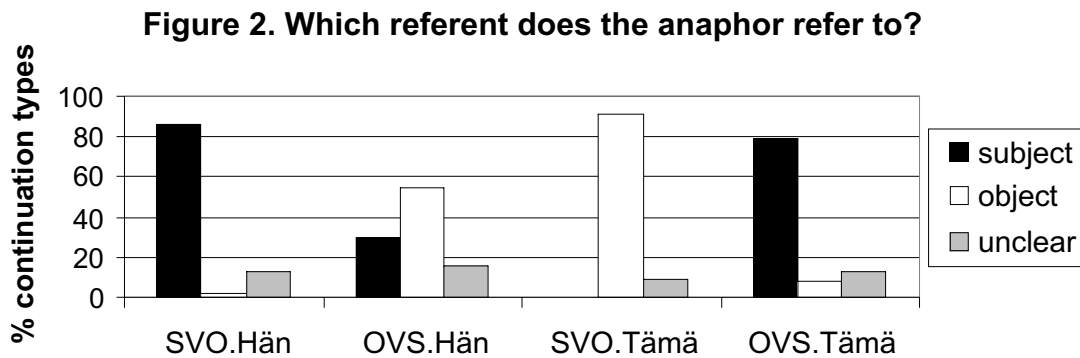
4.2 Experiment 2

This experiment addresses the question whether overt encoding of contextual oldness by means of a pronoun impacts the referential properties of *hän* and *tämä*. Given the findings discussed above, combined with the claim that discourse status affects salience (e.g. Strube & Hahn 1996), one might expect that overt encoding of discourse status by means of a pronoun strengthens this effect. Experiment 2 investigates this possibility. In this experiment, each item consisted of two sentences and the first word of the third sentence (*hän/tämä*) (ex. 9). The first sentence introduces a referent, which is referred to with a pronoun at the beginning of the second sentence. The second sentence (SVO/OVS) has a new referent as its last word.

(9)   [SVO.Hän]
        Puiston penkillä lepäsi lääkäri.   Hän   onnitteli
        Park's bench-on rested doctor-NOM. S/he-NOM congratulated

        opiskelijaa.   Hän .....
        student-PART. S/he-NOM…

        'On the park bench rested a doctor. S/he was congratulating a student. S/he….'

Again, there were four conditions: [SVO.Hän...], [OVS.Hän...], [SVO.Tämä...], and [OVS.Tämä...]. Thirty-two native Finnish speakers participated in this experiment, and each participant wrote continuations for 38 items (8 critical items and 30 fillers) in random order.

The results of the continuations reveal, again, that *tämä* and *hän* are affected differently by word order (see Figure 2). As in Experiment 1, there are significant effects of anaphor type and word order on reference to both subjects and objects (p's<0.05). In other words, (i) the type of anaphoric expression (*hän* vs *tämä*) and (ii) word order (SVO vs. OVS) have a significant effect on whether a particular anaphoric element is interpreted as referring to the preceding subject or object. There is also a significant interaction. (see Kaiser (in preparation) for details.)

Figure 2. Which referent does the anaphor refer to?

Let us now look at the different conditions in more detail. As we saw in Experiments 1 and 1b, *tämä* 'this' tends to refer to the entity (subject or object) introduced at the end of the second sentence – i.e. to the entity most recently introduced into the discussion. In Experiment 2, in the [SVO.Tämä] condition, we see 58/64 (90.6%) object-interpretations, and in the [OVS.Tämä] condition, there are 51/64 (79.9%) subject-interpretations. Thus, just as we saw in Experiment 1 and even more clearly in Experiment 1b, the primary factor that *tämä* is sensitive to is word order.

In contrast, the referential properties of *hän* 's/he' in Experiment 2 do not appear to match what we saw earlier. Now, word order (which is correlated here with pronominalization) has an impact. In SVO order, the most likely antecedent for *hän* is the preceding (pronominal) subject, i.e. the entity that has already been mentioned ([SVO.Hän] = 55/64 (85.9%) subject-interpretations). In OVS order, the referents of *hän* are split between the subject and object. There are 19/64 (29.7%) subject-interpretations, and 35/64 (54.7%) object-interpretations. The preference for the object is significant (t(31)=-3.25, p<.01).

Let us turn to the implications of these results. For *tämä*, the crucial factor in Experiment 2 is word order. We can attribute the strengthening of the word order factor to a preference for *tämä* to refer to clearly discourse-new referents over discourse-old, pronominalized, salient referents. These results thus provide further support for the idea that *tämä* refers to entities that are low in salience.

The results for *hän* are more complex. In Experiments 1 and 1b, we observed that *hän* referred to the subject regardless of grammatical role. However, in Experiment 2, *hän* shows a preference to refer to the pronominalized antecedent in preceding sentence – which is not always the subject. In other words, in both [OVS.Hän] and [SVO.Hän] conditions, the most likely referent of the pronoun is the preverbal, pronominalized referent. However, as the graph shows, there is clearly a difference between [OVS.Hän] and [SVO.Hän]: the preference to refer to the preverbal, pronominalized referent is greater in the [SVO.Hän] condition than in the

[OVS.Hän] condition. This shows that *hän* still has some of the sensitivity to grammatical role that we saw in the earlier experiments.

On the whole, then, Experiment 2 reveals that other factors, beyond grammatical role, play a role in influencing what *hän* refers to – in particular, a pronoun is likely to refer to a preceding, pronominalized, preverbal referent, even if this referent is not the subject. In fact, the idea that the NP form of potential antecedents influences the referential properties of a pronoun is briefly discussed in Beaver (to appear) (see also Ariel 1990:24), and this is an idea that would clearly benefit from further crosslinguistic empirical research.

5. Discussion and conclusions

In light of the results discussed above, we can conclude that *hän* and *tämä* are not mirror images of one another. This suggests that we shouldn't aim to define their referential properties in terms of a single unified notion of salience. Instead, we need to explore how different factors – such as word order, grammatical role, and the form of the antecedent (e.g. full NP vs. pronoun) – are relevant for different referential expressions.

In future work, I plan to look more closely at how the referential form of potential antecedents impacts the referential properties of pronouns and demonstratives, ideally by manipulating both pronominalization and word order. Another interesting area for future research concerns dialects of spoken Finnish, whose referential systems often differ from that of standard Finnish. In the most common spoken dialect, used primarily in the urban areas of southern Finland, the non-human pronoun *se* 'it' is used for human referents, but the pronoun *hän* 's/he' and the demonstrative *tämä* 'this' are also used (see e.g. Seppänen 1998). This situation raises interesting questions, in particular concerning the division of labor of *se* and *hän*. This is a question that clearly merits further work.

On the whole, the results presented in this paper have interesting implications for our understanding of how referential systems work. One possible option is that the system 'assigns jobs' to the elements, such that the functions of one element are fully dependent on the functions of other elements present in the paradigm. This type of approach seems to be implicit in the accessibility hierarchies which suggest that null pronouns are used for more accessible referents than pronouns, which in turn are used for more accessible referents than demonstratives, and so on. Another option is that the different elements can also have properties of their own, independent of the system. The results discussed seem to favor the second option, as they reveal the differences in the factors to which *hän* and *tämä* are sensitive.

References:

Ariel, Mira. 1990. *Accessing NP Antecedents*, Routledge, Croom Helm Linguistics Series, London.

Arnold, Jennifer E. 1998. *Reference form and discourse patterns,* unpublished Ph.D. dissertation, Stanford University, Stanford, CA.

Beaver, David. To appear. 'The Optimization of Discourse Anaphora,' to appear in *Linguistics and Philosophy*.

Brennan, Susan E., Friedman, Marilyn W. & Pollard, Carl J. 1987. 'A Centering Approach to Pronouns,' *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, Stanford, CA, 155-162.

Chesterman, Andrew. 1991. *On definiteness,* Cambridge University Press, Cambridge, England.

Choi, Hye-Won. 1996. *Optimizing Structure in Context: Scrambling and Information Structure,* Ph.D. thesis, Stanford University.

Crawley, Rosalind A. & Stevenson, Rosemary J. 1990. 'Reference in Single Sentences and in Texts,' *Journal of Psycholinguistic Research,* 19(3), 191-210.

Gernsbacher, Morton Ann. 1990. *Language Comprehension as Structure Building,* Lawrence Erlbaum, Hillsdale, NJ.

Givón, Talmy. 1984. *Syntax: A functional-typological introduction*, John Benjamins, Amsterdam/ Philadelphia.

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

Gundel, Jeanette K., Hedberg, Nancy & Zacharski, Ron. 1993. 'Cognitive Status and the Form of Referring Expressions in Discourse,' *Language* 69, 274-307.

Halmari, Helena. 1994. 'On Accessibility and Coreference,' *Nordic Journal of Linguistics* 17, 35-59.

Hoffman, Beryl. 1998. 'Word Order, Information Structure and Centering in Turkish,' in Marilyn A. Walker, Aravind K. Joshi, Ellen F. Prince (eds.), *Centering Theory in Discourse*, Clarendon Press, Oxford, pp. 251-272.

Kaiser, Elsi. In preparation. *Ways of referring and patterns of reference: Anaphoric reference in Finnish.* Ms., University of Pennsylvania.

Kaiser, Elsi. 2000. 'Pronouns and demonstratives in Finnish: Indicators of Referent Salience,' in Paul Baker, Andrew Hardie, Tony McEnery & Anna Siewierska (eds.), *Proceedings of the Discourse Anaphora and Reference Resolution Conference (DAARC 2000)*, University Center for Computer Corpus Research on Language (UCREL) technical papers volume 12, 20-27.

Kalliokoski, Jyrki. 1991. 'Emphathy as motivation for style shifting in narrative,' in Jef Verschueren (ed.), *Levels of linguistic adaptation. Selected papers of the International Pragmatics Conference,* Antwerp, August 17-22, 1987, Vol II, John Benjamins, Amsterdam, pp. 147-161.

Laury, Ritva. 1997. *Demonstratives in Interaction,* John Benjamins, Philadelphia/ Amsterdam.

Lenerz, Jürgen. 1977. *Zur Abfolge nominaler Satzglieder im Deutschen,* Gunter Narr, Tübingen.

Linna, Väinö. 1954/1999. *Tuntematon Sotilas*, WSOY, Helsinki.

Matthews, Alison & Chodorow, Martin S. 1988. 'Pronoun resolution in two-clause sentences: Effects of ambiguity, antecedent location, and depth of embedding,' *Journal of Memory and Language* 27, 245-260.

Prince, Ellen F. 1992. 'The ZPG letter: subjects, definiteness and information status,' in S. Thompson, W. Mann, (eds.), *Discourse description: diverse analyses of a fund raising text*, John Benjamins, Philadelphia/Amsterdam, pp. 295-325.

Rambow, Owen. 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.

Seppänen, Eeva-Leena. 1998. *Läsnäolon pronomit,* Suomalaisen Kirjallisuuden Seura, Helsinki.

Stevenson, Rosemary J., Crawley, Rosalind & Kleiman, D. 1994. 'Thematic roles, focus and the representation of events,' *Language and Cognitive Processes* 9, 519-548.

Strube, Michael. & Hahn, Udo. 1996. 'Functional Centering,' *Proceedings of ACL '96*, 270-277.

Strube, Michael. 1998. 'Never look back: An alternative to Centering,' *Proceedings of Coling-ACL '98*, 1251-1257.

Sulkala, Helena & Karjalainen, Merja. 1992 *Finnish*. Routledge, London.

Tetreault, Joel. submitted. *A Corpus-Based Evaluation of Centering and Pronoun Resolution.*

Turan, Ümit Deniz. 1998. 'Ranking Forward-Looking Centers in Turkish,' in Marilyn A. Walker, Aravind K. Joshi and Ellen F. Prince (eds.), *Centering Theory in Discourse*, Clarendon Press, Oxford, pp. 139-160.

Varteva, Annukka. 1998. 'Pronominit hän ja tämä tekstissä,' *Virittäjä* 2/1998, 202-223.

Vilkuna, Maria. 1995. 'Discourse Configurationality in Finnish,' in Kiss, K. (ed.), *Discourse Configurational Languages*, Oxford University Press, New York, pp. 244-268.

Vilppula, Matti. 1989. 'Havaintoja hän- ja he-prominien käytöstä suomen murteissa,' *Virittäjä* 93, 389-399.