EFFECTS OF TOPIC AND FOCUS ON SALIENCE[*]

Elsi Kaiser,
University of Southern California

elsi.kaiser@usc.edu

Abstract

This paper investigates what factors make a particular referent a good antecedent for subsequent pronominal reference. In particular, it explores two seemingly conflicting claims in the literature regarding the effects of topicality and focusing on referent salience. In light of new experimental results combined with a review of existing work, I conclude that neither topicality nor focusing alone can explain referent salience as indicated by patterns of pronoun reference. Rather, the data provide support for a multiple-factor model of salience (e.g. Arnold 1999). More specifically, the results show that grammatical role has a striking effect: being a subject makes a referent more salient than either pronominalization/givenness or focusing alone. Furthermore, the results of the experiment suggest that the likelihood of subsequent pronominal reference is also influenced by structural focusing and pronominalization, but not as strongly as by subjecthood. I argue that these data are best captured by a multiple-factor model in which factors differ in how influential they are relative to one another, i.e. how heavily weighted they are. A single-factor system does not seem adequate for these data.

1   Introduction

The notion of 'salience' plays a crucial role in theories of reference resolution, as it is widely assumed that the most reduced (and least semantically informative) referring expressions refer to the most salient referents – i.e., the referents which are most prominent, most accessible at that point in the discourse. This, of course, raises the crucial question of what makes a referent salient. A number of factors have been proposed in the literature, and this paper focuses on two apparently contradictory claims, namely that both topicality and focusing – which are often thought of as opposites – increase referent salience. In light of new experimental results combined with a review of existing work, I conclude that neither topicality nor focusing alone can explain referent salience as indicated by patterns of pronoun reference. Rather, the data provides support for a multiple-factor model of salience, suggesting that a referent's salience depends on a number of competing factors which differ in the strength of their influence (see Arnold 1998, 1999).

The structure of this paper is as follows. Section 2 provides an overview of the notion of salience, including claims that have been made in the literature regarding the connections between salience and referential form, and discusses four factors that have been argued to influence the salience of entities, namely subjecthood, givenness, pronominalization and focus. In Section 3 we turn to existing research on the question of whether topical or focused entities are more salient, and Section 4 outlines the open questions that this paper aims to tackle. Section 5 presents the results of the sentence completion experiment, and conclusions and wider implications are discussed in Section 6.

2   Salience

Many researchers have claimed that there are correlations between different kinds of referential expressions (full NPs, pronouns, demonstratives etc) and the level of salience/accessibility of their antecedents (e.g. Gundel, Hedberg and Zacharski 1993, Givón 1983 and Ariel 1990). The general consensus appears to be that the more reduced an anaphoric expression is, the more salient / accessible its antecedent has to be. In this research, the term 'salient' is generally used to mean entities that are currently at the center of attention, i.e. those that are most prominent at that point in the discourse. The view that most researchers assume is summed up in this quote from Arnold (1998): "Loosely speaking, all researchers have observed 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).

However, in order for the claim that salient referents are referred to with reduced anaphoric forms to be meaningful, the notion of salience needs to be defined. More specifically, if we accept the claim that the most salient entities are referred to with the most reduced forms, then we can use pronouns as a tool to investigate the notion of salience in more detail. In other words, we can probe what factors make an entity likely to be referred back to with a pronoun, and assume that these factors are what influence salience.[1] A number of factors have been put forth in the literature as increasing the likelihood of subsequent pronominalization (see Arnold 1998 for an overview), many of which could be regarded as increasing the topicality of a referent. These include occupying the grammatical position of subject, being given information and being realized as a pronoun.

However, before we go any further, it is worth pointing out that the term 'topic' is used differently by different researchers. Strawson (1964) defines the topic of an utterance as "what is of current interest or concern" (Strawson 1964:104). Reinhart (1982) defines the topic of a sentence as "the expression whose referent the sentence is about" (Reinhart 1982:5). Gundel (1985) characterizes topics in terms of 'shared knowledge': "the topic of a speech act will normally be some entity that is already familiar to both speaker and addressee" (Gundel 1982: 92). In more recent work, Prince (2003) and Beaver (2004) use the term 'topic' to refer to the backward-looking center in Centering Theory (Grosz, Joshi & Weinstein 1995); a use which links topicality with pronominalization and givenness. As will become clearer later, in this paper my aim is not to provide an exhaustive definition of topicality; rather, I would simply like to point out that many of the factors that have been claimed to influence referent salience (and which I try to 'pull apart' in order to see what their individual contributions are) have also been linked to the general notion of topicality.

---

[1] The assumption that degree of salience and degree of 'reduction' of the referential form are related is not entirely unproblematic (see Kaiser 2003, Kaiser & Trueswell in press). In particular, it seems that not all referential forms (e.g. pronouns vs. demonstratives in languages like Finnish that allow both to have human antecedents) are sensitive to the same supposedly salience-influencing factors, which argues for a more complex mapping between referential forms and degree of salience of the antecedent that is normally assumed (for details, see Kaiser 2003, Kaiser & Trueswell in press). However, in this paper we are focusing only on one referential form, namely personal pronouns in English, and thus the conclusions should be interpreted as restricted to this form. For the purposes of this paper, we assume that English personal pronouns can be used as a tool to probe referent salience. Even if one wants to argue that this assumption is problematic (e.g. due to the nature of the mapping between salience and referential forms, or due to differences in bottom-up and top-down processing), the results are still relevant: Even if one wants to argue that they do not shed light on the factors that influence salience *per se*, they still shed light on the factors that influence pronoun interpretation (e.g. whether pronoun interpretation is sensitive to only one factor or several differently-weighted factors) as well as the processes that underlie reference resolution.

In contrast to the claims that topicality-related factors make a particular referent especially salient and hence a good antecedent for a pronoun, some researchers have claimed that focusing is what makes a referent salient. In this section, we briefly review these two claims, which seem to conflict, at least at first glance. We will first consider claims regarding subjecthood, givenness and pronominalization, which could be regarded as being related to topicality, and then move onto a discussion of the claims regarding the effects of focusing.

2.1   Subjecthood

A number of researchers have claimed that grammatical role is correlated with salience; more specifically, that entities realized in subject position are more salient than those in non-subject positions (Brennan, Friedman & Pollard 1987, Matthews & Chodorow 1988, Crawley & Stevenson 1990, Stevenson et al. 1994, and McDonald & MacWhinney 1995, *inter alia*). Both corpus evidence and psycholinguistic research support this claim. For example, in a sentence completion study, Crawley & Stevenson (1990) found that when given sentence fragments such as "Shaun led Ben along the path and he….", participants tended to continue the sentence such that the pronoun referred back to the preceding subject more often than to the object. These findings are corroborated by self-paced reading studies, such as Gordon, Grosz and Gilliom (1993) and Stevenson & Urbanowicz (1995), which also found that the grammatical role of an antecedent influences reading times for subsequent pronouns.

2.2   Givenness

Another factor that has been claimed to increase the salience of a referent is givenness, i.e. being 'old' information. For example, Strube & Hahn (1996) argue that the salience of referents is determined by "the functional information structure (IS) of the utterance" (Strube & Hahn 1996:272); more specifically, that "any *context-bound* expression…is given the highest preference as a potential antecedent of an anaphoric or elliptical expression" (Strube & Hahn 1996:272). In other words, when a sentence with a discourse-old referent and a discourse-new referent is followed by an anaphoric expression, the anaphor refers to the discourse-old referent. In related work, Ballantyne (2004) conducted a corpus study of Yapese (Oceanic language in Micronesia) and found that givenness is a better way of ranking referents (in Centering-theoretic terms, leads to more coherent transitions between utterances) than grammatical role or linear order.

2.3   Pronominalization

A number of researchers have found that the referential form with which an entity is realized can affect that entity's salience. Kameyama (1999) claims that a pronominalized referent in non-subject position gains in salience by virtue of being pronominalized, and becomes so salient that it 'competes' in salience with a non-pronominalized entity in subject position. Similarly, Beaver (2004) suggests an Optimality-theoretic approach to anaphora resolution that includes a constraint called SALIENT FORM, which states that "If in the previous sentence discourse entity α was realized by a more minimal form than discourse entity β, then α is more salient than β" (Beaver 2004:31). It is important to note that the constraint SALIENT FORM is different from the idea that the most salient referents are referred to with the most reduced forms, since, as Beaver point out, SALIENT FORM "implies that being pronominalized makes a referent salient *in the future*" (Beaver 2004:31 fn 30, italics added, see also Kehler 2001:169).

As mentioned above, factors such as subjecthood, givenness and pronominalization could all be regarded as increasing the topicality of a referent. Thus, one might be tempted to conclude

that a salient referent is a topical referent. However, as we will see in the next section, not all researchers agree that salience is necessarily linked to topicality. In addition, not all researchers agree that a conglomeration of factors is what determines salience. Some researchers seem to either assume or claim that salience is determined by only one factor (e.g. see Strube & Hahn 1999), whereas others argue in favor of a multiple-factor view (e.g. Ariel 1990, Arnold 1998, see also Givón 1983). We return to this question in Section 4.

2.4   Focus

This section reviews claims that focused referents are more salient than non-focused referents. Like the term 'topic', the term 'focus' would also benefit from some clarification. Focus is often divided into contrastive focus (or identificational focus, to use Kiss's (1998) term) and presentational focus (information focus, according to Kiss). The existing psycholinguistic work investigating focusing has tended to look at the effects of contrastive focus, since it has used structures such as it-clefts ('It was Mary who called Lisa'), which are usually regarded as expressing contrastive (identificational) focus.[2] The experiment, described in Section 5, uses both clefts and *in-situ* focus constructions, but due to the context in which they occur, both involve contrastive focus. (Green and Jaggar (2003) claim that *in-situ* focus can also be interpreted contrastively.) Thus, the claims made in this paper regarding focus only apply to contrastive focus. The effects of presentational focus are an important direction for future work (see also Hajičová, Kubon & Kubon 1992).

Now, let us turn to the research that supports the claim that (contrastively) focused entities are more salient than non-focused ones. In a cognitive psychology experiment, Hornby (1974) presented participants with pictures and sentences, and asked people to say whether the sentence matches the picture. When participants were presented cleft sentences (e.g. 'It is the girl who is riding the bicycle'), Hornby found that the participants were better at detecting mismatches when the mismatching information was focused than when it was presupposed. This suggests that participants attend more to the non-presupposed, focused part of the sentence (but see Delin 1990). In related work, Singer (1976) probed people's memory of focused and non-focused referents using sentences such as 'It was the king who led the troops' and 'It was the troops that the king led.' He found that focused referents are remembered better than non-focused referents. Thus, it seems that focused information is noticed and remembered better than non-focused information, which could be regarded as result of its being perceived or represented differently from non-focused information due to its being more salient.

Extending this work to reference resolution, Almor (1999) conducted a reading time study which found that reference to focused referents is read faster (i.e., presumably processed with greater ease) than reference to non-focused referents. Almor tested sentence pairs such as those in (1) (with focused subjects) and (2) (with focused objects), and found that (1a) is read faster than (1b), and (2b) is read faster than (2a), suggesting that an anaphoric expression (e.g. the bird, the fruit) is interpreted faster when it refers to an antecedent that is in focus than when it refers to an antecedent that is not focused (in this case, presupposed).

(1) a.   It was the robin$_{focus}$ that ate the apple.
          The bird seemed very satisfied.

---

[2] It is important to note that corpus studies have shown that regarding all clefts as structures where the clefted constituent is contrastively focused and the rest of the sentence is presupposed is a gross oversimplification (see e.g. Delin 1990). However, the it-clefts used in the experiment described in Section 5 were all very simple in that the focused constituent was new information and the rest of the sentence was given (see example (8)).

    b. It was the robin$_{focus}$ that ate the apple.
       The fruit was already half rotten.
(2) a. What the robin ate was the apple$_{focus}$.
       The bird seemed very satisfied.
    b. What the robin ate was the apple$_{focus}$.
       The fruit was already half rotten.

If we assume that a referring expression referring to a highly salient referent is read faster (processed with greater ease) than one referring to a low-salience referent, these results indicate that the focus of a cleft is more salient than the non-focus. These findings also appear to be compatible with the claim that clefts involve low topic continuity (Givón 1983), which can be roughly paraphrased as stating that a cleft is used when the discourse is shifting to a new center of attention, namely the entity that is focused in the cleft.

In a different tradition, Hajičová, Kubon & Kubon (1992) claim, on the basis of computationally-oriented corpus work on Czech, that entities in the focal part of an utterance are the most salient, and entities in the topical part are less salient. In contrast to the other research on focus, however, Hajičová et al. use the term 'focus' to refer not to the focus of clefts, but to the 'contextually non-bound' parts of an utterance, i.e. those parts which are, roughly, new information. Moreover, it is worth noting that according to Hajičová et al., certain pronominal forms tend to refer to focused entities and others prefer topical entities – i.e. even though they explicitly claim that focused referents are more salient than topical referents, they do not claim that the most reduced referring expressions refer to the most salient (i.e. focused) referents.

Thus, there exists an intriguing division in the literature. On the one hand, a number of factors have been claimed to render a particular referent highly salient and thus a good antecedent for a subsequent anaphor, and many of these factors could also be argued to be related to the general notion of topicality. On the other hand, it has also been argued that (contrastively) focused referents are especially salient and prominent in people's mental models of the discourse. In the next section, we turn to some existing experimental work that aims to shed light on this seeming contradiction.

3   Topic vs. focus: Which is more salient?

In light of the contrasting claims presented in the preceding sections, let us now turn to existing experimental work (Arnold 1999, Cowles 2003) that aims to resolve the conflicting claims regarding the salience of topics and the salience of foci.

Arnold (1999) conducted a number of psycholinguistic experiments investigating the salience of topics and foci. She used pronouns as a tool for probing which referent in the preceding discourse is the most salient. To test whether topical and focused referents are more salient than other referents, she tested three-sentence 'mini-narratives' such as (3) and (4) in a rating study. She manipulated whether the second sentence was clefted[3] or not, whether the subject of the third sentence referred to the first- or second-mentioned character of the second sentence ((c) vs. (c')), and whether the subject of the third sentence was a pronoun or a name. The subscripts on the examples illustrate which constituents Arnold assumes to be topics and

---

[3] Arnold used clefts with 'the one' rather than it-clefts or wh-clefts, but she notes that the 'one'-construction has been called a cleft with a lexical head by Prince (1978). In Arnold's cleft sentences, strictly speaking, the subject and object of the matrix copular sentence (e.g. 'the one [+ relative clause] was Emily') both refer to the same entity, namely the focus. The topic in these one-clefts is the subject of the relative clause modifying 'one' (e.g. 'the one [**he** decided on at last] was Kysha'). This is in contrast to the it-clefts used in the experiment described in this paper, where the topic is either the matrix subject or the matrix object.

foci. It is important to note that she constructed these sentences such that the non-clefted ones contained a topic (she follows existing research in assuming that the subject functions as a topic) but no syntactically marked focus, and the clefted ones contained a clear focus but no strong topic. This was done because the aim of this experiment was to investigate topics and foci independently of each other.

(3) a. The guests were nervously standing around in the living room, trying to decide which person to talk to.
   b. Ann$_{topic}$ decided to say hi to Emily[4] first. (NON-CLEFT)
   c. Emily/She looked like the friendliest person in the group.
   c'. Ann/She hated to be in a room full of people where no-one was talking.

(4) a. The guests were nervously standing around in the living room, trying to decide which person to talk to.
   b. The one Ann[5] decided to say hi to first was Emily$_{focus}$. (CLEFT)
   c. Emily/She looked like the friendliest person in the group.
   c'. Ann/She hated to be in a room full of people where no-one was talking.

The results of the rating study indicate that, in general, topics and foci are more salient than other referents. More specifically, with nonclefted sentences (3), participants prefer pronouns in the third sentence to refer to the subject of the second sentence (the topic), as in (3c'), and with clefted sentences (4), participants prefer pronouns to refer to the object of the second sentence (the focus), as in (4c). Furthermore, for referring to entities other than the topic in non-clefts and the focus in clefts, full names are preferred.

To investigate what happens when topics and foci are directly pitted against each other, Arnold conducted a production study, where participants were given sequences of sentences such as those in (5), ending in either a clefted or an non-clefted sentence ((5a) vs. (5b)). In this experiment, Arnold established a particular referent as the discourse topic by introducing it in the subject position of the first sentences, and referring back to it with a pronoun in the next two sentences. The participants' task was to provide a continuation for the story.

(5)   Ron was looking through his address book, trying to make up his mind.
      He had an extra ticket to the opera, but he didn't know which friend to invite.
   a. The one he$_{topic}$ decided on at last was Kysha$_{focus}$ / Fred$_{focus}$. (CLEFTED)
   b. At last he$_{topic}$ decided on Kysha$_{focus}$ / Fred$_{focus}$. (NON-CLEFTED)

An analysis of the pronouns occurring in participants' continuations reveals a strong preference to use pronouns to refer to the subject of the preceding sentence (the topic), regardless of whether the sentence was clefted (97%) or not clefted (98%). This suggests that topics are more salient than foci, regardless of the syntactic form of the sentence.

In a third experiment, Arnold investigated the difference between a well-established discourse topic (as in (5), a referent that has already been mentioned in preceding discourse by the time it is realized, as a pronoun, as the topic of a cleft) and a sentence topic (as in (4), a referent that is realized, as a full NP, in the topic position of a cleft but had not been mentioned in the preceding discourse). The results of a rating study indicate that when a clear discourse topic exists, participants prefer the focus of the cleft to be referred to with a name, but when no discourse topic is present, there is no such preference. Arnold concludes that this is because the absence of a clear discourse topic makes it possible for the focus to be relatively more salient than it could be in the presence of an overwhelmingly salient discourse topic; in other

---

[4] As Arnold notes, this referent is realized as an oblique object, and not syntactically marked for focus.

[5] Note that this referent is realized as an embedded subject, which Arnold notes is not highly topical. It is also new information.

words, she regards salience as a competitive phenomenon, where "where the representations of different referents in a particular discourse compete for activation" (Arnold 1999:28).

Using a different methodology, Cowles (2003) reaches a somewhat different conclusion. She uses cross-modal priming to investigate the impact of discourse-topicality, sentence-topicality and contrastive focus on referent salience. She uses clefts to mark focus, and uses the term 'discourse topic' for referents that have been realized twice in subject position, and the term 'sentence topic' for referents that have been realized once in subject position. The results indicate that "[a]ll three information statuses [discourse topic, sentence topic and contrastive focus, EK] appear to make their referent more likely to be interpreted as the antecedent of a subsequent pronoun" (Cowles 2003:93). In fact, in contrast to Arnold who found that established discourse topics are more salient than foci, Cowles concludes that "two information structure types that are considered distinct …. appear to have the same psychological effect" (2003:94). However, it appears that Cowles tested referents that were subjects whereas Arnold tested subject topics and object foci. It seems that their studies differ not only in methodology but also in the nature of the materials, which may be partly responsible for the different findings.

In sum, although existing experimental work suggests that topics and contrastive foci are more salient than other referents (see also Navarretta 2002), the results conflict when it comes to the question of which is more salient, a topic or a focus.

4   Effects of different factors

As we saw in Section 2, subjecthood is very often regarded as being correlated with salience/topicality, but based on the research discussed in Section 3, it appears that (to the best of my knowledge) existing experiments on pronoun resolution have not fully investigated possible consequences of grammatical role on the effects of topicality and focusing. In light of the claim that grammatical role influences salience, I would like to suggest that in order to improve our understanding of how topicality and focusing influence salience, we should investigate both subject and non-subject topics and foci. For example, how do topics in object position compare to foci that are objects?

Let us briefly consider the nature of the relation between subjects and topicality. As already noted, many researchers have observed that entities realized in subject position tend to be interpreted as topical. However, it seems that this does not always have to be the case. Consider the example below:

(6)   After serving little more than a year in jail, Cruz-Mendoza was deported for a third time in
        January, records and interviews show. U.S. Border Patrol agents arrested him in Arizona a
        month later. At that point, he could have been charged with a felony….
        (Los Angeles Times, October 30, 2005)

Most approaches would agree that the topic of the second sentence is the referent of the object pronoun 'him' and not the subject of the sentence 'U.S. Border Patrol agents' (see e.g. Prince 2003 on Centering Theory, Beaver 2004 and others)[6]. Thus, it is not the case that topics are restricted to occurring in subject position. In the case of foci, it is also clear that they are not restricted to occurring in object position, as illustrated by examples such as (7) below. Here we see a subject it-cleft, where the subject 'Lisa' is in focus:

---

[6] Of course, this statement would not be compatible with a theory where the notion of topic is inherently linked to subjecthood. However, the burden would then be on such a theory to show that 'U.S. Border Patrol agents' is more topical than the referent of 'him.'

(7) Mesmerized, I called them up and asked for an interview to discuss how they cast applicants for reality TV. First, I talked to Debbie, who said she'd get back to me. Instead, it was Lisa who returned my call.
(www.themorningnews.org/archives/manufacturing_reality/mirror_mirror.php)

If we combine the observation that topics do not have to be subjects and foci do not have to be objects with the well-known claim that subjecthood influences referent salience, it becomes clear that investigating subject topics and object foci, for example, may result in overestimation of the effects of topicality as a result of associating it with subjecthood. However, looking only at subject foci and subject topics may also be insufficient, since if subjects turn out to be highly salient simply due to their subject status, then this could potentially 'wash out' effects of the topic/focus distinction. One of the main aims of the experiment described in this paper is to investigate subject-topics, subject-foci, object-topics and object-foci (in both clefted and nunclefted sentences) in order to see which factors are the most influential in determining which referents are good antecedents for subsequent pronouns.

More generally, these issues are related to the larger question of how different factors interact during reference resolution. In particular, as mentioned earlier, is it the case that a single factor determines which entities can be referred to with pronouns in subsequent discourse, or might it be the case that a number of factors, perhaps with different degrees of influence (different weights) all play a role? In other words, if we accept the claim that the most salient entities are referred to with the most reduced referential forms, then we can use pronouns as a tool to ask: Does one unique factor determine salience, or can multiple factors interact? If multiple factors interact, are they all weighted equally, or are some more influential than others? These are the questions that the experiment in the next section explores.

5   Experiment

In order to shed light on the issues sketched out above, this experiment investigates how subjecthood, pronominalization, semantic focusing and syntactic focusing influence subsequent pronoun use. The specific aim of the experiment is to pull apart the subject-topic / object-focus correlation that is common in previous experimental work. On a more general level, in disassociating these factors the experiment will also help us to better understand the issues sketched out above regarding the interaction and degree of influence of different kinds of information during reference resolution.

We manipulated syntactic form (cleft vs. SVO) and the grammatical role of the topical/focused constituent, as illustrated in example (8). Thus, there are four conditions, which will be referred to with the following shorthand labels: [SVO.Object=focus], [SVO.Subject=focus], [Cleft.Object=focus] and [Cleft.Subject=focus]. The referent marked as 'focus' in (8) is always semantically focused thanks to the context, and in the clefted conditions it is also structurally focused as a result of being in the focus position of the cleft. The referent subscripted as 'topic' in the example in (8) is discourse-old and pronominalized, and follows Prince's (2003) and Beaver's (2004) use the term 'topic' to refer to the Centering Theory notion of backward-looking center. (However, my use of the subscript 'topic' in (8) is not intended to convey the claim that the referent of the pronoun is more salient than the focused expression. See Section 5.1.)

The participants' task was to provide a natural-sounding continuation sentence using the pronoun prompt that followed each critical sentence. They were told to imagine that someone has just made the claim in part A, and that they were now responding to this other person by saying part B and providing a continuation. Participants were recorded using a Tascam digital tape recorder and a Shure unidirectional headmounted microphone.

All verbs were agent-patient verbs, as defined by Stevenson et al (1994). This was done in order to control for any potential verb focusing effects. Both human referents mentioned in the sentences were of the same gender; either both (stereotypically) male or both (stereotypically) female. There were 16 target items and 16 fillers.

(8)   A: The maid scolded the bride.
       a. B: No, that's wrong! She$_{topic}$ scolded the secretary$_{focus}$. She….
       b. B: No, that's wrong! The secretary$_{focus}$ scolded her$_{topic}$. She…
       c. B: No, that's wrong! It was the secretary$_{focus}$ that she$_{topic}$ scolded. She…
       d. B: No, that's wrong! It was the secretary$_{focus}$ who scolded her$_{topic}$. She…

Participants' (n=24) continuations were digitized and transcribed, and the referent of the prompt pronoun in each of the continuations was double-coded by two coders working independently. Disagreements were resolved by discussion. If it was not clear who the pronoun refers to, the item was coded as 'unclear.' Table 1 provides some examples.

(i) A: The waiter criticized the sailor.
B: No, that's wrong! He criticized the businessman. He gave him too small a tip.
Coded as: he = businessman, i.e. object
(ii) A: The waiter criticized the sailor.
B: No, that's wrong! He criticized the businessman. He didn't get a very good tip.
Coded as: he = waiter, i.e. subject
(iii) A: The maid scolded the bride.
B: No, that's wrong! She scolded the secretary. She told me about it after it happened.
Coded as: she = unclear

Table 1. Coding samples

5.1 Predictions

As mentioned earlier, there are different hypotheses regarding the nature of the relation between factors such as subjecthood and focusing. In particular, some researchers seem to espouse a single-factor view, which assumes that one factor plays a decisive role in determining which referents can be subsequently referred to with pronouns, whereas others appear more supportive of a multiple-factor view.

Let us start by considering the predictions that a single-factor view would make for the factors investigated in this experiment, namely subjecthood, pronominalization, semantic focusing and syntactic focusing.[7] If subjecthood is the one factor that determines referent salience, the prediction is that prompt pronouns will refer to preceding subjects, regardless of NP form or topic/focus status. In contrast, if pronominalization (and givenness) determines referent salience, we predict that prompt pronouns will refer to whatever is pronominalized in the preceding sentence, regardless of whether it is the subject or the object, clefted or uncleftled. However, if semantic focusing is the one factor that determines referent salience, prompt

---

[7] These factors are not fully crossed in this design – partly due to the nature of the phenomena being investigated. For example, syntactically focused entities are also necessarily semantically focused, but not vice versa. Furthermore, in this design pronominalization and focusing are in complementary distribution in the sense that a particular referent is either pronominalized or focused, but never neither and never both.

pronouns are predicted to refer to the focused referent, regardless of syntactic role or sentence structure (cleft vs SVO). If structural focusing (clefting) alone is what determines referent salience, the prediction is that prompt pronouns will refer to the focus of the cleft in clefted sentences, but it is not clear what the prediction would be for unclefted sentences.

Let us now turn to the multiple-factor view, according to which two or more factors could be influencing referent salience, and hence the likelihood of subsequent pronominal reference. Let us assume, for expository ease, that all four factors are relevant and weighted equally. The rightmost column of Table 2 summarizes which referent in each condition is predicted to be most likely to be referred to with a subsequent pronoun.

In the [SVO.Object=focus] condition (line A of Table 2), two factors (subjecthood and pronominalization) contribute to the salience of the subject. Semantic focusing contributes to the salience of the object. This could also be cast in terms of activation in the participant's mental model of the discourse: both subjecthood and pronominalization increase the level of activation of the subject, and semantic focusing increases the activation level of the object. Thus, if all factors are weighted equally, the subject 'wins' over the object; it is more activated. In the [SVO.Subject=focus] condition (line B), pronominalization points towards the object, but subjecthood and semantic focusing both point towards the subject. Thus, if all factors are weighted equally, we again predict that the subject wins out over the object. Note that in this condition the subject is focused, whereas in the preceding condition it was the discourse-old, pronominalized referent.

Now, turning to the first of the two cleft conditions, in the [Cleft.Object=focus] condition (line C) we see that both subjecthood and pronominalization increase the salience of the subject, but structural and semantic focusing both point towards the object. Thus, in contrast to the [SVO.Object=focus] condition, now the object is focused both structurally and semantically. Assuming that this would have a stronger effect than semantic focusing alone (see also Navarreta 2002 on the effect of information-structural devices being used to mark focus) leads us to the prediction that in the [Cleft.Object=focus] condition, the subject and object are tied. Put differently, they have equal levels of activation. Finally, in the [Cleft.Subject=focus] condition (line D), everything except pronominalization is pointing towards the subject: subjecthood status, structural focusing and semantic focusing. This leads to the prediction that the subject has a higher level of activation than the object.

It is important to note that I have been assuming that all factors are weighted equally; i.e. that they make equal contributions to the salience levels of the subject or the object. Of course, this might very well not turn out to be the case. In fact, in constraint-based models of language processing (e.g. MacDonald et al., 1994, Tanenhaus & Trueswell, 1995, Trueswell et al., 1994, see also Bates and MacWhinney, 1989) which claim that language processing is guided by weighted constraints, the constraints differ in their weights and hence can differ in magnitude of the impact they have on language processing, depending also on the number of competing alternatives (see also Arnold 1998).

		Subject	Pronom	Sem foc	Str foc	Overall
A	She scolded the SECRETARY.	S	S	O	…	S (top)
B	The SECRETARY scolded her.	S	O	S	…	S (foc)
C	It was the SECRETARY that she scolded.	S	S	O	O	?
D	It was the SECRETARY who scolded her.	S	O	S	S	S (foc)

Table 2. Multiple-factor view (if each factor is weighed equally)

5.2   Results and discussion

Overall, participants' continuations reveal an overall preference to interpret prompt pronouns as referring to subjects, regardless of whether the subject was a topic or a focus. This is illustrated in Figure 1, which shows the average subject advantage scores for the four conditions. These scores were calculated by taking the proportion of subject continuations in each condition and subtracting from that the proportion of object continuations. Thus, a positive subject advantage score indicates more subject continuations than object continuations, and a negative subject advantage score indicates more object continuations than subject continuations. As Figure 1 clearly shows, the subject advantage score is positive in all four conditions, indicating that there were more subject continuations than object continuations. Participants were more likely to interpret the subsequent pronoun as referring to the subject than to the object.

The overall subject preference indicates that subjecthood matters more than pronominalization, more than semantic or structural focusing. It seems that subjecthood makes both topics and foci good antecedents for a subsequent pronoun.
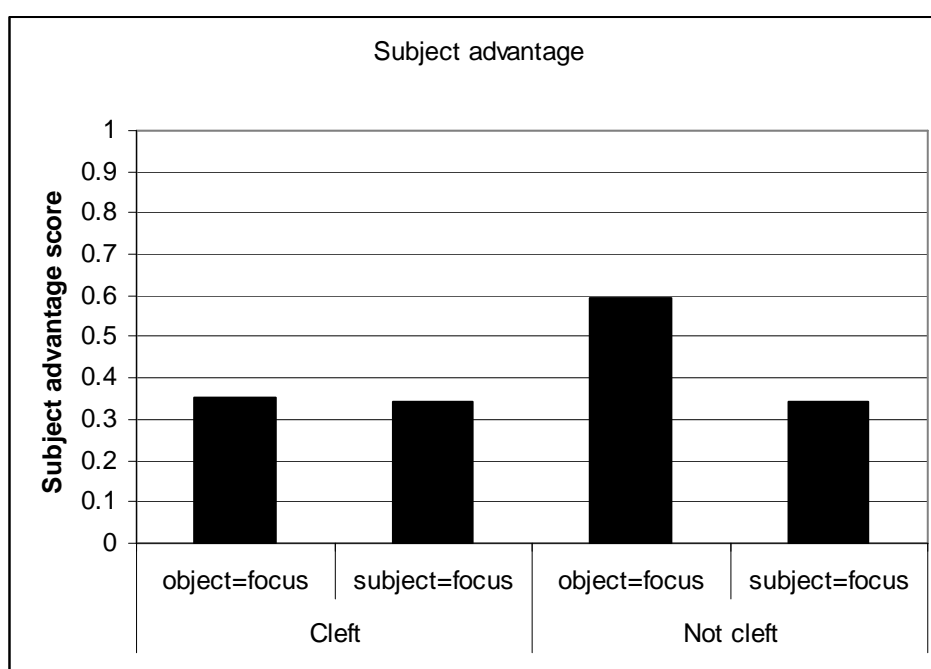

Figure 1. Subject advantage scores (proportion of subject continuations minus proportion of object continuations)

However, let us now look more closely at the different conditions. Let us first compare conditions with clefted and non-clefted focused objects. As Figure 1 shows, in the conditions with focused objects, there is a greater subject advantage in the non-clefted condition [SVO.Object=focus] than in the clefted condition [Cleft.Object=focus]. Why is this? A possible reason for the weaker subject preference in the [Cleft.Object=focus] condition is that in this particular condition, both semantic and structural focusing point towards the object. In other words, the only difference between the [SVO.Object=focus] and the [Cleft.Object=focus] conditions is structural; the latter is clefted (see also Navarreta 2002 for related corpus work on clefts in Danish). Thus, the difference between these conditions can be straightforwardly captured if one assumes, as Arnold (1999) suggests, that salience is a competitive phenomenon. More specifically, I hypothesize that the combination of syntactic and semantic focusing increases the salience of the object sufficiently so that it can compete

with the subject and weaken the subject advantage in the clefted version. Thus, the results indicate that clefting a focused object increases its chances of being referred to by a subsequent pronoun, compared to a non-clefted focused object.[8]

However, if this is the reason for the difference that emerges between the [SVO.Object=focus] and [Cleft.Object=focus] conditions, then why do the [SVO.Subject=focus] and [Cleft.Subject=focus] conditions not show as strong a subject preference as the [SVO.Object=focus] condition does? Why do they both show a subject preference comparable to that in the [Cleft.Object=focus] condition, as Figure 1 clearly illustrates? In the two Subject=focus conditions, there is no structurally focused object to pull participants away from the subject, so that cannot be the reason for the weakened subject preference. However, it is important to note that in the Subject=focus conditions, subjecthood and pronominalization are pitted against each other. As we saw in Section 2, in previous work both of these factors have been found to influence referent salience. Again, if we assume that salience is a competitive phenomenon, then it follows that the conflict between subjecthood and pronominalization is responsible for the weaker subject preference we see in the Subject=focus conditions, since pronominalization increases the salience (or activation) of the object, which leads to it being better able to compete with the subject.

It is worth noting that in these particular conditions, the SVO vs. cleft distinction does not appear to have any effect on the strength of the subject advantage; it is not the case that the [Cleft.Subject=focus] condition has a stronger subject preference than the [SVO.Subject=focus] condition. However, in light of the claim that subjecthood is more heavily weighted than structural focusing, this is not entirely surprising, as it could be explained simply by the much greater influence of subjecthood masking or 'swamping' the effects of structural focusing. In other words, structural focusing seems to have a stronger effect on the salience of objects than on (already 'inherently' salient) subjects.

Taken as a whole, the results support the multiple-factor model. However, it is clearly not the case that all factors are weighted equally. The results suggest that subjecthood is more influential (weighted more heavily) than either pronominalization or structural or semantic focusing. However, the effects of subjecthood are modulated by structural focusing and pronominalization. As we saw, the contrast between the [SVO.Object=focus] and the [Cleft.Object=focus] conditions suggests that structural focusing can increase the salience of a referent. Furthermore, the finding that the subject advantage is stronger in the [SVO.Object=focus] condition than in the [SVO.Subject=focus] and [Cleft.Subject=focus] conditions suggests that if pronominalization and subjecthood are pitted against each other, the effects of subjecthood are weakened. In sum, even though the pattern of results is fairly complex and will of course need to be investigated more in future work, it seems clear that we are dealing with a competition-based system sensitive to multiple factors which are weighted differently.

6   Conclusions

Let us now return to the conflict sketched out at the beginning of this paper, namely the seemingly contradictory claims that topics are the most salient or that foci are the most

---

[8] One might also wonder whether parallelism is at work here. According the Smyth's (1994) parallelism account, pronouns prefer antecedents that are in the same syntactic position as the pronoun itself. However, this preference only holds, according to Smyth, when the relevant sentences both have the same global constituent structure and the thematic roles of the verbs in the two sentences match. Consider, for example, a sentence like 'Peter hit John. Alex pinched him.' However, an examination of participants' continuations suggests that the required degree of matching across sentences does not seem to be consistently present. This casts doubt on the idea that parallelism is at work here.

salient. The results of the sentence completion experiment suggest that both claims are partly right, but that the picture is actually more complex and involves multiple interacting factors mediating between referents competing for salience. Crucially, the results show that subjecthood makes both topics (pronominalized, discourse old referents) and foci good antecedents for a subsequent pronoun. The observation that being a subject makes a referent more salient than pronominalization / givenness or focusing alone suggests that looking only at subject topics and object foci may result in an inadvertent overestimation of the effects of topicality, and that looking only at subject topics and subject foci may not be very fruitful due to the overwhelming effects of subjecthood.

However, subjecthood is not the only thing that matters; there are also effects of structural focusing and pronominalization. As mentioned in Section 5, the subject advantage difference between sentences with clefted and unclefted focused objects suggests that structurally clefting a focused object influences its salience – but to a lesser degree than subjecthood. Furthermore, we also saw in Section 5 that the finding that the [SVO.Subject=focus] and [Cleft.Subject=focus] conditions do not show as strong a subject preference as the [SVO.Object=focus] can be straightforwardly explained if we assume that pronominalization increases the salience of a referent. Like structural focusing, pronominizalition has an effect on the salience of a referent, but is not as 'powerful' a factor as subjecthood.

Of course, many questions still remain open, and further research is needed to investigate the validity of the hypotheses presented here, both in English and other languages. For example, given data suggesting that different factors are weighed differently, I would like to know more about the reasons or causes of these weight differences, as well as the extent of crosslinguistic variation in this domain. In future work, I would also like to investigate the intonational patterns used in these kinds of contexts, in particular in the *in-situ* focus sentences as compared to the clefts, in order to see how prosodic information is contributing to the reference resolution process. The distinction between stressed and unstressed pronouns is also a crucial question for future work. Another issue that would benefit from further research is the relation between agentivity and subjecthood. This experiment only investigated agentive subjects, and thus confounds agentivity and subjecthood. Comparing agentive and non-agentive subjects (e.g. experiencers) would shed light on the question of whether it is the structural notion of subjecthood or the semantics of agentivity that is behind the subjecthood effect observed in the sentence completion experiment.

In sum, the results of the experiment presented here suggest that in order to begin to untangle the seemingly conflicting claims regarding the impact of topicality and focusing on salience, subjecthood must be taken into account. Furthermore, the results indicate that the strong effect of subjecthood on referent salience is modulated by effects of pronominalization and structural focusing. Thus, as a whole, the data presented here are best captured by a multiple-factor model in which factors differ in how influential they are relative to one another, i.e. how heavily weighted they are, and referents compete for activation (see Arnold 1998, 1999, *inter alia*).[9] A single-factor system does not seem adequate for this kind of data, and thus it seems reasonable to conclude that salience (at least insofar as we are measuring salience by looking at likelihood of subsequent pronominal reference) is not a monolithic concept.

---

[9] It would also be very interesting to see whether the findings reported here could be captured in an Optimality-theoretic system, perhaps similar to the one in Beaver (2004).

References

Almor, A. (1999) Noun-phrase anaphora and focus: The informational load hypothesis. Psychological Review, 106(4), 748-765.

Ariel, M. 1990. *Accessing NP antecedents*. London: Routledge, Croom Helm.

Arnold, J. 1998. *Reference form and discourse patterns.* Ph.D. dissertation, Stanford University.

Arnold, J. 1999. *Marking Salience: The similarity of topic and focus*. Manuscript, Univ. of Pennsylvania.

Ballantyne, H.G. 2004. Givenness as a Ranking Criterion in Centering Theory: Evidence from Yapese. *Oceanic Linguistics*, Volume 43(1): 49-72.

Bates, E. & MacWhinney, B. (1989). *Functionalism and the competition model*. In E. Bates, & B. MacWhinney (Eds.), The crosslinguistic study of sentence processing. Cambridge: Cambridge University Press.

Beaver, D. 2004. The optimization of discourse anaphora. *Linguistics and Philosophy* 27(1):3-56.

Brennan, Susan E., Friedman, Marilyn A. & Pollard, Charles J. 1987. A Centering approach to pronouns. *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, Stanford, Calif., 155-162.

Cowles, H.W. 2003. *Processing Information Structure: Evidence from Comprehension and Production.* Ph.D. dissertation, UCSD.

Crawley, R.J. and Stevenson, R.J. 1990. Reference in single sentences and in texts. *Journal of Psycholinguistic Research* 19(3):191-210.

Delin, J. 1990. A psychological investigation of the processing of cleft constructions. Blue Book note no. 2.

Givón, T. 1983. *Topic continuity in discourse: A quantitative cross-language study.* Amsterdam: John Benjamins.

Gordon, P.C., Grosz, B.J., & Gilliom, L.A. 1993. Pronouns, names, and the Centering of attention in discourse. *Cognitive Science* 17:311-347

Green, M. and Jaggar, P. 2003. Ex-situ and In-situ Focus in Hausa: syntax, semantics and discourse. In Lecarme, J (Ed.) 2003. *Research in Afroasiatic Grammar II.* [CILT 241]. Amsterdam: John Benjamins: 187-213.

Grosz, B., Joshi, A. & Weinstein, S. 1995. Centering: A Framework for Modelling the Local Coherence of Discourse. *Computational Linguistics*, 2(21).

Gundel, Jeanette K., Hedberg, Nancy & Zacharski, Ron. 1993. Cognitive status and the form of referring Expressions in Discourse. *Language* 69, 274-307.

Gundel, Jeanette. 1974. *The role of topic and comment in linguistic theory.* University of Texas Ph.D. dissertation.

Hajicova, E., Kubon, V., & Kubon, P. 1992. *Stock of shared knowledge - a tool for solving* pronominal anaphora. Proc. of COLING-92, Vol. 1, 127–133.

Hornby, P.A. 1974. Surface structure and presupposition. *Journal of Verbal Learning and Verbal Behavior* 13:530-538.

Kaiser, E. 2003. *The quest for a referent: A crosslinguistic look at reference resolutio*n. Ph.D. dissertation, University of Pennsylvania.

Kaiser, E. & J.C. Trueswell (In press) Investigating the interpretation of pronouns and demonstratives in Finnish: Going beyond salience. To appear in E. Gibson & N. Pearlmutter (eds), *The processing and acquisition of reference.* Cambridge, Mass.: MIT Press.

Kameyama, M. 1999. Stressed and unstressed pronouns: Complementary preferences. In *Focus: Linguistic, Cognitive, and Computational Perspectives,* ed. P. Bosch & R. van der Sandt, 306-321. Cambridge University Press.

Kehler, A. 2001. Coherence, Reference and the Theory of Grammar. Stanford: CSLI

Kiss, K. 1998. Identificational focus vs. Information Focus. *Language* 74(2):245-273.

MacDonald, M.C., N.J. Pearlmutter & M.S. Seidenberg. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101: 676-703.

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

McDonald, J.L. & MacWhinney, B. 1995. The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language* 34:543-566.

Navarretta, C. 2002. Combining Information Structure and Centering-based models of salience for resolving intersential pronominal anaphora. In A. Branco, T. McEnery and R. Mitkov (eds), *Proceedings of the 4th Discourse Anaphora and Anaphor Resolution Colloquium*, pp.135-140. Lisbon: Edições Colibri

Prince, E.F. 1978. A Comparison of Wh-Clefts and It-Clefts in Discourse. *Language* 544:883-906.

Prince, E.F. 2003. *The Yiddish impersonal pronoun men 'one' in discourse.* Presented at the Annual Meeting of the Linguistic Society of America, Atlanta, January 2003. (available online at www.ling.upenn.edu/~ellen)

Reinhart, Tanya. 1982. *Pragmatics and linguistics: an analysis of sentence topics.* University of Indiana Linguistics Club. (also *Philosopica* 1981, 27.53-94).

Singer, M. 1976. Thematic structure and the integration of linguistic information. *Journal of Verbal Learning and Verbal Behavior* 15:549-558.

Smyth, R. 1994. Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research*, 23:197-229.

Stevenson, R. and Urbanowicz, A. 1995. Structural focusing, thematic role focusing and the comprehension of pronouns. In *Proceedings of the 17ᵗʰ Annual Conference of the Cognitive Science Society*, 328-332. Pittsburg, PA.

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

Strawson, P.F. 1964. Identifying reference and truth-values. *Theoria* 30:96-118. Reprinted in *Semantics: An Interdisciplinary Reader in Linguistics, Philosophy and Psychology* , ed. by Danny D. Steinberg and Leon A. Jakobovits. 1971. Cambridge: Cambridge University Press.

Strube, M. & Hahn, U. 1996. Functional Centering. In *Proceedings of ACL '96*, 270-277.

Strube, M. & Hahn, U. 1999. Functional Centering: Grounding Referential Coherence in Information Structure. *Computational Linguistics* 25(3).

Tanenhaus, M. K., & Trueswell, J. C. 1995. *Sentence comprehension*. In J. L. Miller & P. D. Eimas (Eds.), Handbook of Perception and Cognition: Volume 11, 217-262. Academic Press, San Diego.

Trueswell, J.C., Tanenhaus, M.K., & Garnsey, S. 1994. Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language* 33:285-318.