Pronouns Without Explicit Antecedents: How do We Know When a Pronoun is Referential?

Jeanette K. Gundel*, Nancy Hedberg†, Ron Zacharski**

*University of Minnesota, Minneapolis, USA and NTNU, Trondheim, Norway
  gunde003@umn.edu
†Simon Fraser University, Burnaby, Canada
  hedberg@sfu.ca
** New Mexico State University, Las Cruces, USA
  raz@crl.nmsu.edu

Abstract

Pronouns without explicit noun phrase antecedents pose a problem for any theory of reference resolution. We report here on an empirical study of such pronouns in the Santa Barbara Corpus of Spoken American English, a corpus of spontaneous, casual conversation. Analysis of 2,046 third person personal pronouns in fourteen transcripts indicates that 326 (or 16%) lack NP antecedents. These pronouns fall into a variety of subtypes. 89 refer to entities that are inferrable from an activated frame or script, or are otherwise easily accommodated. In 129 cases, "it" could refer to a fact, proposition event, activity, situation, or reason which has been evoked by a previous non-NP. 80 cases of it were classified as pleonastic. In this paper we focus on some interesting subclasses of pronouns which could be analyzed as either referring to entities of various degrees of abstractness that were introduced by or implied in previous discourse, or as non-referential or pleonastic pronouns. Such cases include possible truncated clefts, inferential constructions, possible truncated extraposition constructions, and certain non-specific uses of they.
1. Introduction

The referent of prototypical pronoun has been recently introduced into the discourse by a noun phrase, i.e. the pronoun has an explicit noun phrase antecedent, as in (1):

(1) a. My neighbor's Bull Mastiff bit a girl on a bike.
   b. It's the same dog that bit Mary Ben last summer. (Gundel et al., 2000)

It is well known, however, that such an antecedent is neither necessary nor sufficient for appropriate pronoun use, especially in the case of personal pronouns. Examples like (2) show that use of it is not always acceptable even when there is a recent NP antecedent:

(2) a. Sears delivered new siding to my neighbors with the Bull Mastiff.
   b. #It's/That's the same dog that bit Mary Ben last summer. (Gundel et al., 1993)

Moreover, an NP antecedent, or an explicit antecedent of any sort, is not always necessary, a fact which poses problems for any theory of reference resolution (see Cornish 1999, Gundel, Hedberg and Zacharski 2000, Byron 2000, inter alia).

We report here on an empirical study of such pronouns in the Santa Barbara Corpus of Spoken American English, a corpus of spontaneous, casual conversation. Particular attention will be paid to cases where it isn't clear whether the pronoun refers to a specific fact, situation, proposition, etc. or whether it is the subject of a truncated cleft (Hedberg 2000), a truncated extraposition construction, or is otherwise possibly non-referential.

2. Quantitative study

As mentioned above, we obtained our data from the Santa Barbara Corpus of Spoken American English Part-1 (DuBois, et al., 2000). We analyzed the first 790 intonation units in each of 14 transcripts, ranging from 681 to 1379 seconds [RON CHECK THIS ON YOUR TRANSCRIPTS]. The transcripts represent conversations between two to four speakers in a variety of settings. [AGAIN, CHECK THIS, RON, FOR YOUR TRANSCRIPTS.]

2.1. Coding

We coded each of the 2,046 third-person personal pronouns (excluding false starts) in the corpus for whether or not they had an NP antecedent, and if not, then what type of thing they referred to. We divided the examples into different classes of referents, distinguishing between inferrables, facts, propositions, activities, events, situations, and reasons. We also coded pronouns as pleonastic. We now turn to a brief introduction of each of these categories.

2.1.1. Inferrables

In (3), the referent of the pronoun is an inferrable (Prince 1981) or an indirect anaphor (Gundel et al., 2000). The referent of the pronoun hasn't been explicitly mentioned in the discourse. Rather, the addressee has to infer from mention of the kids across the street that she refers to their mother.

(3) [Talking about how the kids across the street threw paint in their yard.] Those kids are just — And she's pregnant with another one. (2.294)

There were a variety of relations between the antecedent trigger (Cornish 1999) and the inferrable pronoun, for example specific entity to generic kind, generic kind to set of specific entities, individual to couple, or individual to group (c.f. the poset relations of Hirschberg 1991). It is more common in the data, however, for there not to be an explicit antecedent trigger; rather the referent of the pronoun is inferred from discussion of a particular situation, such as filing a police report or a class, to they referring to the police or he referring to the teacher. In many of these cases, it is plausible to assume that the referent is inferred from an activated frame or script (see Gundel, et al., 2000 for more discussion of inferrables). Moreover, as will be discussed in Section 3 below, some inferrables might not be referential at all.

MAYBE SOMETHING ABOUT INFERRABLES MORE OFTEN BEING ENCODED WITH DEFINITE DESCRIPTIONS, AND ABOUT ACCOMMODATION. ALSO MAYBE SOME EXAMPLES OF THE FRAME EXAMPLES SHOULD BE GIVEN.

2.1.2. Facts and propositions

In (4), the speaker is saying that the fact every horse's hoof is shaped different doesn't matter.

(4) You go look, and every horse's hoof is shaped different. It doesn't matter. (1.347)

Propositions are the object of belief and other propositional attitudes, such as doubt. In (5), the speaker is saying she couldn't believe the proposition that the woman is going to be a ferrier.

(5) And she's gonna be a ferrier.. I couldn't believe it. (1.802)

MAYBE SOMETHING ABOUT ASHER AND GUNDEL, HEGARTY AND BORTHEN. IT WOULD BE NICE TO ORGANIZE THIS SECTION AND THE FOLLOWING ONE AROUND THE DEGREE OF WORLD IMMANENCE OR SOMETHING. IT IS INTERESTING THAT THERE ARE VERY FEW PROPOSITIONS, FACTS AND REASONS, AND MORE ACTIVITIES, EVENTS AND SITUATIONS REFERRED TO WITH A PERSONAL PRONOUN. IT WOULD BE NICE IF YOU WOULD LOOK AT RON'S AND MY DATA AND TRY TO HELP US CLASSIFY EVENTS VS. SITUATIONS SINCE SITUATIONS ARE SUPPOSED TO HAVE LESS WORLD IMMANENCE

2.1. Coding

We coded each of the 2,046 third-person personal pronouns (excluding false starts) in the corpus for whether or not they had an NP antecedent, and if not, then what type of thing they referred to. We divided the examples into different classes of referents, distinguishing between inferrables, facts, propositions, activities, events, situations, and reasons. We also coded pronouns as pleonastic. We now turn to a brief introduction of each of these categories.

2.1.1. Inferrables

In (3), the referent of the pronoun is an inferrable (Prince 1981) or an indirect anaphor (Gundel et al., 2000). The referent of the pronoun hasn't been explicitly mentioned in the discourse. Rather, the addressee has to infer from mention of the kids across the street that she refers to their mother.

(3) [Talking about how the kids across the street threw paint in their yard.] Those kids are just — And she's pregnant with another one. (2.294)

There were a variety of relations between the antecedent trigger (Cornish 1999) and the inferrable pronoun, for example specific entity to generic kind, generic kind to set of specific entities, individual to couple, or individual to group (c.f. the poset relations of Hirschberg 1991). It is more common in the data, however, for there not to be an explicit antecedent trigger; rather the referent of the pronoun is inferred from discussion of a particular situation, such as filing a police report or a class, to they referring to the police or he referring to the teacher. In many of these cases, it is plausible to assume that the referent is inferred from an activated frame or script (see Gundel, et al., 2000 for more discussion of inferrables). Moreover, as will be discussed in Section 3 below, some inferrables might not be referential at all.

MAYBE SOMETHING ABOUT INFERRABLES MORE OFTEN BEING ENCODED WITH DEFINITE DESCRIPTIONS, AND ABOUT ACCOMMODATION. ALSO MAYBE SOME EXAMPLES OF THE FRAME EXAMPLES SHOULD BE GIVEN.

2.1.2. Facts and propositions

In (4), the speaker is saying that the fact every horse's hoof is shaped different doesn't matter.

(4) You go look, and every horse's hoof is shaped different. It doesn't matter. (1.347)

Propositions are the object of belief and other propositional attitudes, such as doubt. In (5), the speaker is saying she couldn't believe the proposition that the woman is going to be a ferrier.

(5) And she's gonna be a ferrier.. I couldn't believe it. (1.802)

MAYBE SOMETHING ABOUT ASHER AND GUNDEL, HEGARTY AND BORTHEN. IT WOULD BE NICE TO ORGANIZE THIS SECTION AND THE FOLLOWING ONE AROUND THE DEGREE OF WORLD IMMANENCE OR SOMETHING. IT IS INTERESTING THAT THERE ARE VERY FEW PROPOSITIONS, FACTS AND REASONS, AND MORE ACTIVITIES, EVENTS AND SITUATIONS REFERRED TO WITH A PERSONAL PRONOUN. IT WOULD BE NICE IF YOU WOULD LOOK AT RON'S AND MY DATA AND TRY TO HELP US CLASSIFY EVENTS VS. SITUATIONS SINCE SITUATIONS ARE SUPPOSED TO HAVE LESS WORLD IMMANENCE.
THAN EVENTS. MAYBE SOME OF OUR SITUATIONS ARE COMPLEX EVENTS.

2.1.3. Activities, Events, Situations and Reasons

In (6), the speaker is saying that the activity of doing the translations needs to worth her time.

(6) I'm going to do some translations for her and stuff? And um, you know, I have to make at least 50 dollars or so, to make it worth my time. (7.323)

In (7), it is the event of the speaker having thrown down the addressee's blouse that he thought was funny.

(7) I threw a green pepper down your blouse. (SNIF) .. I thought it was funny. (3.384).

Events differ from activities in that their referent is evoked by a whole prior sentence, not just a verb phrase. The distinction is not always a clear one.

In (8), the speaker is saying that she was into the situation of the car thief showing television viewers how not to get their car broken into.

(8) He's gonna show us, you know, how not to protect your car, not to get it, you know, ripped off man. Cause, you know, I -- , yeah, I was into it. (6.31)

Situations differ from events in being less specific. The situation referred to by the pronoun in (6) is evoked by a wide expanse of prior discourse. Again the distinction is not always clear.

Finally, there was one case of a pronoun that we classified as referring to a reason, namely (9):

(9) A. Guess kids bones, jut like .. grow back really fast.
B. Mhm.
C: Yeah. I think they're really soft to start with.
A. They're made of rubber. That's it.

Here, the speaker is saying that kids' bones being made of rubber is the reason for them growing back really fast. As will be discussed in Section 3 below, some of the pronouns that we coded as non-referential pleonastic pronouns could instead be interpreted as referring to reasons, namely certain subjects of truncated cleft sentences.

2.1.4. Pleonastics

We coded for four types of pleonastic pronouns: cleft pronouns, inferential pronouns (Delahunty, 2001) extraposition pronouns and atmospheric pronouns, as exemplified in (10)-(13), respectively:

(10) Was it Trich who told me she was pregnant? (13.216)

(11) It's just it's a whole different earth. (3.738)

(12) I just think it's so damn weird we're here. (5.529)

(13) It rained during the dry season. (3.526)

In Section 3 below, we discuss three types of cases in which it is not clear whether a pronoun is pleonastic or refers to a higher-order entity such as a reason, cause, situation or event.

2.2. Results

The results of our corpus study are shown in Table 1, which presents the total number of pronouns with each type of referent, if any, and also shows a few cases that weren't included in the primary classification because they were exophoric, included in an idiom such as God damn it or had indeterminate reference.

<table>
<thead>
<tr>
<th>Type</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP antecedent</td>
<td>1720</td>
<td>84.07</td>
</tr>
<tr>
<td>Inferrable</td>
<td>89</td>
<td>4.35</td>
</tr>
<tr>
<td>Fact</td>
<td>7</td>
<td>0.34</td>
</tr>
<tr>
<td>Proposition</td>
<td>6</td>
<td>0.29</td>
</tr>
<tr>
<td>Activity</td>
<td>28</td>
<td>1.37</td>
</tr>
<tr>
<td>Event</td>
<td>12</td>
<td>0.59</td>
</tr>
<tr>
<td>Situation</td>
<td>75</td>
<td>3.67</td>
</tr>
<tr>
<td>Reason</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>Pleonastic</td>
<td>80</td>
<td>3.91</td>
</tr>
<tr>
<td>Idiom</td>
<td>9</td>
<td>0.44</td>
</tr>
<tr>
<td>Exophoric</td>
<td>9</td>
<td>0.44</td>
</tr>
<tr>
<td>Indeterminate</td>
<td>6</td>
<td>0.29</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>2046</strong></td>
<td><strong>100.00</strong></td>
</tr>
</tbody>
</table>

Table 1: Pronouns in Santa Barbara Corpus

326 of the third person personal pronouns in the corpus, or 16%, lacked NP antecedents. This figure can be compared to the percentages of pronouns without NP antecedents reported on in Byron (2002). She reports that Eckert and Strube (2000) found that 22% of pronouns in a set of Switchboard dialogs had non-NP antecedents and that 33% lacked antecedents altogether; Byron and Allen (1998) found that 50% of pronouns in the TRAINS corpus lacked NP antecedents; and that Botley (1996) found that 20% of pronouns in his corpus lacked NP antecedents. In at least some of these cases, it can be assumed that demonstrative pronouns were included in the study, so the total number of pronouns lacking NP antecedents can be expected to be higher than the figure we found since it is well known that demonstrative rather than personal pronouns are often used to refer to higher-order entities such as propositions, situations, and speech acts. REFERENCE WEBBER; GUNDEL, HEGGARTY AND BORTHEIN PAPERS?.

3. Determining referentiality

We were particularly interested in examples of pronoun use that were difficult to classify as referential or non-referential. There were four types of such pronouns.

3.1. Vague inferrables
In (14) and (15), they seems to refer to people in general and doesn't have a specific referent.

(14) And they say that if there's six years between children, there's not that much rivalry. (7.1247)

(15) And they probably didn't have to wash their salads back then, because they didn't know what was on them. (3.165)

Byron (2000) suggests that such uses of they are type-identifiable but not referential on the Givenness Hierarchy of Gundel et al. (1993), and refer to some unspecified group of people. Gundel et al. (2000) point out that sentences like (10) can be replaced by agentless passives with no loss of information content:

In (16), the phrase they had could be replaced by there was and the pronoun seems even more nonreferential than the type in (14) and (15).

(16) He said I didn't get done working until after nine.

Cause that five-car pile up they had between Hardin and Crow? (7.414)

There seems to be a continuum between clearly referential inferrables and nonreferential inferrables as reference proceeds, for example, from a specific woman across the street; to the people in the office, whoever they might be; to people in general.

3.2. Truncated cleft pronouns

In (17), the sentence in B's reply is a truncated cleft. The cleft clause who stole Hector's radio has been elided.

(17) A: It's obvious now that this guy w-- -- This was the one who stole . Hector's [radio]. I mean, .. nobody came out and told you, guess what, I confess.

B: Oh, we knew. .. We knew. .. We figured it had to be Michael. (2.70)

Hedberg (2000) presents a theory of cleft sentences that claims that cleft pronouns combine with the cleft clause to form a discontinuous definite description. Cognitive-pragmatic conditions of use for this description are determined by where the description falls on the Givenness Hierarchy of Gundel, et al. (1993), with the cleft pronoun functioning as a determinant in a nontruncated cleft and as a pronoun in a truncated cleft. Thus the cleft clause material must be at least uniquely identifiable in a nontruncated it-cleft, familiar in a nontruncated that-cleft, activated in a nontruncated this-cleft, and in focus in a truncated it-cleft.

In example (17) the material is in focus because the conversation has been about a car thief appearing on a television show who turns out to be Michael and also about the theft of Hector's radio. The one who stole Hector's radio has just been introduced in the previous contribution.

Hedberg claims that the cleft pronoun is not a meaningless, pleonastic element but rather is referential in the sense that a determiner is in the case of full clefts, and in the sense that at least some pronouns are in the case of truncated clefts. To understand in what sense the subject pronoun of a truncated cleft is referential, compare Hedberg's example in (18) with example (17).

(18) My heart beat fast, for I had thought that as the discoverer of the body I would be the first to be called; but to my surprise, it was Marcel.

Here, the truncated cleft could be replaced by a full cleft, It was Marcel who was called. Hedberg's claim is that the subject pronoun in the truncated cleft co-refers with the first to be called.

In both (17) and (18), the antecedent for the pronoun has been introduced as the complement of an inverted pseudocleft. Thus, both (17) and (18) can be paraphrased as pseudoclefts: We figured the one who stole Hector's radio had to be Michael; but to my surprise, the one who was called was Marcel. The sentences here analyzed as truncated clefts could be analyzed, then, as pseudoclefts with pronominalized subjects, and the sense in which the subject is referential is the same sense as the free relative subject of a pseudocleft is referential.

HOW REFERENTIAL IS THIS, THEN? IT DOESN'T SEEM TO HAVE THE SAME DEGREE OF WORLD IMMANENCE AS AN ORDINARY INDIVIDUAL ENTITY. BUT THE NEXT PARAGRAPH IS SUPPOSED TO INTRODUCE EXAMPLES IN WHICH THE PRONOUN COULD REFER TO A REASON OR A CAUSE, WHICH DO SEEM TO HAVE SOME FAIRLY HIGH DEGREE OF WORLD IMMANENCE.

There are several examples of truncated clefts in the Santa Barbara corpus whose subjects we classified as pleonastics in Table 1 but which could equally well be classified as referring to a reason or a cause evoked in prior discourse. For example, (19) and (20):

(19) A: So that's why you're interested in death?

B: Maybe it's because my parents were old? When I was young? (5.499)

(20) A: What do you think makes em look African?

B: Their mustaches?

A: Is it the way their little beard goes? (2.517)

Here again, the truncated cleft could be replaced by a full cleft: Maybe it's because my parents were old that I am so interested in death; Is it the way their little beard goes that makes em look African? Furthermore, the sentences could be paraphrased as pseudoclefts: Maybe why I'm so interested in death is because my parents were old; Is what makes em look African the way their little beard goes? However, the cleft pronoun in both (19) and (20) can also be taken as referring to a reason or cause evoked in the previous question.

In (21), the pronoun seems to refer to a cause, but it is not introduced in a question or in an inverted pseudocleft clause. Instead the existence of a cause can be inferred from the fact that the speedometer fell, and the cause is referred to with the pronoun. The full cleft paraphrase here would be I knew exactly what it was that caused it.
The nonleft paragraph would be *I knew exactly what the cause was.*

(21) I saw my .. my speedometer just go Brr=. .. like that just dow=n,. You know, and I knew exactly what it was.

**There is something not entirely referential about these superscriptional (Higgins 1973) phrases, the pseudocleft subjects with free relatives and phrases like the reason here. The speaker is going to be filling in a value for the variable and so these phrases are value free (Barwise and Perry). Should I explain this? I guess so.**

### 3.3. Inferential subjects

Delahunty (2001) analyzes sentences like (22) as inferential constructions.

(22) *It s just == it gets dust accumulated in it, see it s all over the TV=. (1.815)*

Inferentials consist of a finite clause embedded as complements of a pleonastic copular matrix clause. It is frequently the case that the matrix clause is modified by the particle *just* in English. Delahunty claims that positive inferentials such as the one in (22) characterize the proposition expressed in the subordinate clause as unlikely to be considered but they assert its relevance. The proposition may be interpreted as an implicated premise or conclusion, and in the former case, as an explanation, reason, or cause; and in the latter case, as a result, consequence, or conclusion.

In (22) the proposition is interpreted as the cause of the some substance blowing out of the air conditioner. It does appropriate to interpret the inferential proposition as relevant but unlikely to be considered. Although Delahunty explicitly argues that the subject pronoun in an inferential sentence is pleonastic, it could be interpreted as referring to a cause, reason, result, consequence, etc. implied in previous discourse.

I have to check Delahunty s article to find out what he really says. I m relying here on a previous article on Spanish inferentials and on what I remember from his Berlin paper. I d like to see what he has to say about just as well, if anything.

### 3.4. Truncated extraposition pronouns

Finally, we classified xx pronouns as pleonastic extraposition construction subjects, for example, those shown in (23) and (24).

(23) But, for me it s really difficult to pick up a book about death. (5.217)

(24) And finally it dawns on Lisabeth that she doesn t see Mom that much. (6.275)

In (23) the extraposed element is an infinitival clause, and the sentence could be paraphrased, *But, for me, to pick up a book about death is really difficult.* And in (24) the extraposed clause is a finite clause, and the sentence can be paraphrased, *And finally that she doesn t see Mom that much dawns on Lisabeth.* Infinitival and finite clauses were the most frequent types of extraposed elements in our data, but we also found some gerund clauses and some free relative clauses.

Extraposition construction subjects are generally analyzed as pleonastic and we classified them as such in Table 1. But what s interesting is that there are some *it* subjects that we didn t classify as pleonastic but which could be analyzed as the subject of truncated extraposition sentences, for example the sentences in (24) and (25):

(24) And the second week they were just like (YELL), and so I had to scream at them, all week long. And it was really awful, cause I felt horrible about it. (4.83)

(25) You can t really tell when they blush. *It s very unusual.* (4.298)

Both of these sentences can be paraphrased as extraposition constructions: *And it was really awful that I had to scream at them, all week long; It s very unusual for them to blush.* But the subject pronouns could equally well be analyzed, as we analyzed them, to refer to higher-order entities evoked in previous discourse: a situation in the case of (24), or a generic event or process in the case of (25).

### 4. Conclusion

### 5. References


WE STILL NEED ABOUT TWO PAGES AND THESE PAGES ARE LONG. WHAT DO YOU THINK WE CAN ADD? WE CAN ADD MORE EXAMPLES, PRESUMABLY, BUT I D LIKE MORE SUBSTANCE THAN THAT. ANY IDEAS YOU HAVE WOULD BE MUCH APPRECIATED.