Justified Sloppiness in Anaphoric Reference

Massimo Poesio,[†] Uwe Reyle,[‡] and Rosemary Stevenson[*]

[†] *University of Essex*
Department of Computer Science
poesio@essex.ac.uk

[‡] *Universität Stuttgart*
Institut für Maschinelle Sprachverarbeitung
uwe@ims.uni-stuttgart.de

[*] *University of Durham*
Department of Psychology

Abstract. A corpus of spoken dialogues was analyzed to identify cases in which the addressee of an utterance containing an anaphoric pronoun does not appear to have enough evidence to resolve that pronoun, yet doesn't appear to find the pronominal use infelicitous. The two patterns of anaphoric use we found that fit these conditions suggest three conditions under which *justified sloppiness* in anaphoric references is not perceived as infelicitous. Preliminary controlled experiments indicate that subjects do find anaphoric pronouns that satisfy the justified sloppiness conditions significantly easier to process than pronouns occurring in minimally different contexts in which these conditions are not satisfied.

Keywords: Underspecification, pronoun resolution

1. Motivations

> *The reason why I hate critics ... is that they write sentences like this:*
>
> > *.... Flaubert does not build up his characters, as did Balzac, by objective cultural description; in fact, so careless is he of their outward appearance that on one occasion he gives Emma brown eyes (14); on another, deep black eyes (15); and on another blue eyes (16).*
>
> *I must confess that in all the times I read* Madame Bovary*, I never noticed the heroine's rainbow eyes. Should I have? Would you? Put it another way: IS THERE A PERFECT READER SOMEWHERE, A TOTAL READER?*
>
> (From J. Barnes, *Flaubert's Parrot*, Picador, 1984, p. 74-76)

The quote above expresses an intuition shared by many computational semanticists: namely, that readers and listeners do not seem to always construct complete interpretations of everything they read or hear. The

possibility that utterance meanings may occasionally be 'incomplete' or 'underspecified' may lead to a fundamental rethink of traditional ideas about how meaning is constructed, and has therefore generated a great interest in recent years, testified, e.g., by the collection (van Deemter and Peters, 1996), by a recent issue of the *Journal of Semantics* on this topic, and by a number of workshops. Computational semanticists, primarily concerned with lexical and scopal underspecification, focused on developing logical characterizations of the type of interpretation that may be assigned to an utterance when its meaning remains underspecified. As a result, it has been shown that it is possible to provide a logical characterizations of the complete space of possible interpretations (see, e.g., (Alshawi and Crouch, 1992; van Eijck and Jaspars, 1996; Muskens, 1995; Pinkal, 1995; Poesio, 1991; Poesio, 1996; Reyle, 1993; Reyle, 1996) as well as the papers in (van Deemter and Peters, 1996)).

However, not much empirical evidence has yet been found supporting this intuition, except perhaps for the case of lexical underspecification (Frazier and Rayner, 1990; Copestake and Briscoe, 1995). On the contrary, there is a lot of evidence suggesting that that a number of semantic interpretive processes take place immediately, just like syntactic interpretation does. (Well-known results concerning the incrementality of semantic interpretation are discussed in (Swinney, 1979) for lexical interpretation and (Tanenhaus et al., 1995) for anaphora resolution.) Our long-term goal is to examine the evidence for and against underspecification in anaphora resolution by identifying cases in which anaphoric expressions, and especially pronouns, are not completely interpreted, and determining the interpretation they receive. (We focus on anaphoric interpretation rather than scope disambiguation –the area that has motivated the most work on underspecification in computational semantics (Reyle, 1993; Poesio, 1994; Muskens, 1995; Pinkal, 1995)– because empirical evidence on anaphora resolution is much easier to collect.) Our research combines corpus analysis and more traditional psychological experimentation: using corpus analysis first to identify contexts in which pronouns may remain underspecified, then running controlled psychological experiments to verify whether indeed this is the case.

In this paper we first discuss the results of a study of a corpus of task-oriented spoken conversations that led us to identify contexts in which an apparently ambiguous anaphoric expressions does not seem to result in a problem being signalled by the other conversational participant. We then propose a preliminary hypothesis concerning what these cases have in common, and finally discuss psychological experiments supporting our hypothesis.

2. Underspecification in Reference: Psychological Evidence

Whereas work on underspecification in computational syntax such as (Sturt and Crocker, 1996) is supported to some extent by empirical evidence –e.g., on the greater or lesser facility of certain syntactic reanalysis processes–the work on underspecification in computational semantics has been driven much less by psychological results, since such evidence was, until recently, pretty minimal. There is, however, an increasing number of studies pointing out cases in which aspects of semantic interpretation are not immediately resolved. These studies are discussed in some length in (Poesio, 1999), to appear as (Poesio, 2004); a useful summary of the psycological evidence is in (Sanford and Sturt, 2002).

The best-known results concerning semantic underspecification are the studies by (Frazier and Rayner, 1990). Frazier and Rayner observed that cases of homonymy like *pitcher* or *records* could be experimentally separated from cases of polysemy like *newspaper*: whereas words belonging to the first class would originate garden paths when subsequent context disconfirmed the preferred interpretation at the point the word (e.g., *record*) was encountered, as in (1d), no garden paths were observed for polysemous words like *newspaper* in (2d).

(1) a. *After they were scratched, the records were carefully guarded.*
b. *After the political takeover, the records were carefully guarded.*
c. *The records were carefully guarded after they were scratched.*
d. *The records were carefully guarded after the political takeover.*

(2) a. *Lying in the rain, the newspaper was destroyed.*
b. *Managing advertising so poorly, the newspaper was destroyed.*
c. *Unfortunately the newspaper was destroyed, lying in the rain.*
d. *Unfortunately the newspaper was destroyed, managing advertising so poorly.*

For the case of anaphoric reference, the evidence is mixed. On the one hand, it is known (see, e.g., (Tanenhaus et al., 1995)) that definite descriptions referring to objects in the visual situation are interpreted immediately. It is also known, however, that the case with anaphoric definite descriptions and other anaphoric expressions such as pronouns is more complex. Although the interpretation of pronouns begins immediately, anaphoric pronouns in ambiguous contexts (i.e., with multiple same-gender potential antecedents) remain uninterpreted until the end of the sentence (Gernsbacher and Hargreaves, 1988). And some evidence suggests that pronouns not referring to a focal entity are not

immediately interpreted (Garrod and Sanford, 1985; Garrod et al., 1994). Garrod et al. (1994), for example, tested the effect of gender, focusing, and verb bias using an eye-tracker and materials like those in (3). A context paragraph was used to establish either a male or a female entity as focus, and to introduce a second entity whose gender either matched or didn't match that of the focused entity. Then a target sentence was presented, containing either a masculine or a feminine pronoun, and a verb biased either towards the focused entity (*sank*) or towards the second entity (*jumped*).

(3) A dangerous incident in the pool
Elizabeth$_1$ / **Alexander**$_1$ *was an inexperienced swimmer and wouldn't have gone in if* **the male lifeguard**$_2$ *hadn't been standing by the pool. But as soon as she*$_1$*/he*$_1$ *got out of her*$_1$*/his*$_1$ *depth she*$_1$*/he*$_1$ *started to panic and wave her*$_1$*/his*$_1$ *hands about in a frenzy.*

a. *Within seconds, she sank into the pool.*
b. *Within seconds, he sank into the pool.*
c. *Within seconds, she jumped into the pool.*
d. *Within seconds, he jumped into the pool.*

First-pass reading times for the verb indicated that conflicts between the interpretation of the pronoun and the verb (as in (3c)) were only immediately detected, resulting in a slowed reading time for the verb, when both gender information and focus information converged on the interpretation of the pronoun; otherwise, the conflict was only detected later. According to Garrod *et al.*, this suggested that unless both these conditions were satisfied, the interpretation of the pronoun was delayed.

3. Finding (Anaphoric) Underspecification in Corpora

3.1. Methodology

How is it possible to use a corpus to find cases in which aspects of the meaning of an utterance –in our case, an anaphoric expression–remain underspecified? Our method has been to analyze task-oriented conversations from the TRAINS corpus collected at the University of Rochester, `http://www.cs.rochester.edu/research/speech/dialogues.html`, and to look for cases in which (i) more than one potential antecedent of an anaphoric expression matches it in gender and number, (ii) no focusing principle we are aware of makes one of the interpretations preferred, and yet (iii) the recipient of the utterance is able to accomplish the task without signalling a problem.

The assumption underlying this approach is that in task-oriented conversations, unlike so-called 'cocktail-party' situations, the participants need to signal when they didn't understand something, as S does in 24.5 in in (4):

(4)
```
23.7    M: what would be faster
23.8     : to send
23.9     : an engine
23.10    : from Elmira
23.11    : to
23.12    : ... one of the boxcars
23.13    : or from
23.14    : Avon
24.1    S: well there's
24.2     : there's a boxcar
24.3     : already _at_ Elmira
           [3sec]
24.4     : and
24.5     : t / YOU MEAN TO GO
           TO CORNING
```

We identified several patterns of pronominal use that satisfy these conditions: three annotators (including two of the authors) agree that a use of a pronoun is ambiguous, and the absence of repair signals (such as *Sorry, what did you mean?*, or *I didn't understand*) indicates that the listener didn't seem to have a problem with the expression being ambiguous. We discuss these patterns in the rest of this section, and the conclusions that we drew from these cases in the next.

3.2. References to Mereologies

A first clear pattern emerging from the corpus is illustrated by the following example (from dialogue d91-3.1):[1]

(5)
```
3.1     M: can we .. kindly hook up
3.2      : uh
3.3      : engine E2 to the boxcar at ..
           Elmira
4.1     S: ok
```

[1] The same person plays the system's role in all TRAINS-91 dialogues; 8 different subjects play the manager role, and each of them is involved in two dialogues. The first part of the identification code of the dialogue says which speaker was involved; the second which dialogue it was - for example, d91-1.1 is the first dialogue for speaker 1, d91-1.2 is the second dialogue for speaker 1, d91-2.1 is the first dialogue for speaker 2, etc.

```
5.1     M: +and+ send it to Corning
5.2      : as soon as possible please
6.1     S: okay
          [2sec]
7.1     M: do let me know when it gets
           there
8.1     S: okay it'll /
8.2      : it should get there at 2 AM
9.1     M: great
9.2      : uh can you give the
9.3      : manager at Corning instructions
           that
9.4      : as soon as it arrives
9.5      : it should be filled with
           oranges
10.1    S: okay
10.2     : then we can get that filled
```

In this example, it's not clear whether the pronoun *it* in 5.1 refers to *the engine E2* which has been hooked up to *the boxcar at Elmira*, to the boxcar itself, or indeed whether that matters. It's only at utterance 9.5 that we get evidence that *it* probably referred to *the boxcar at Elmira*, since it is only boxcars that can be filled with oranges; yet, if anything, focusing theories would predict engine E2 to be the antecedent, since engine E2 is the direct object, the THEME, and comes first (Sidner, 1979; Stevenson et al., 1994; Grosz et al., 1995; Poesio and Stevenson, To appear).

Note that this type of pronoun use cannot be viewed as an example of vagueness, at least not according to the standard vagueness tests (Lakoff, 1970; Zwicky and Sadock, 1975). Whereas the ellipsis test, for example, would suggest that *a glove* in (6a) is indeterminate, because it's possible for John to have lost its left glove and for Bill to have lost its right one, the same test applied to *it* in (6b) suggests that this expression is ambiguous, in that it's not possible to interpret *Then, John should check if IT gets to Bath in time, and Bill should too* as meaning that John should check that the engine gets to Bath in time, whereas Bill should check that the boxcar gets there in time:

(6) a. *John lost a glove, and Bill did too*
b. *Let's hook the engine to the boxcar.*
Then, John should check if IT gets to Bath in time, and Bill should too

Another context also apparently leading to unproblematic use of pronouns with multiple matching antecedents is illustrated by (7). In this

example, the pronoun *that* in 27.4 may refer either to the orange juice or to the tanker car in which the orange juice has been loaded:

(7)
```
26.1    S: okay
27.1    M: so then we'll
27.2     : ... we'll be in a position to
27.3     : load the orange juice into the
           tanker car
27.4     : ... and send that off
```

In order to characterize in a more systematic fashion the possible interpretations of *that* in 27.4 in this last example (and of *it* in (5), 5.1) we will borrow some notation from Link (1983). We will write $oj \oplus tc$ to indicate the object that has oj and tc as subparts, and $a \lhd b$ to say that a is a mereological part of b. With this notation, we can formalize the first and most obvious property of examples (5) and (7): namely, that actions like *hooking up* and *loading* are performed that create a new object $a \oplus b$ out of the potential antecedents a and b (e.g., $oj \oplus tc$ in (7)).

The second property of these examples is that four interpretations for the pronominal expression are possible. The complete list of the possible interpretations of *that* in (7), 27.4 is:

$$that = oj,\ tc,\ oj \oplus tc, \text{ or an indeterminate } x \lhd (oj \oplus tc)$$

This latter interpretation $(x \lhd (oj \oplus tc))$ is what has been called a P-UNDERSPECIFIED INTERPRETATION in (Poesio, 1999)–i.e., a 'disjunctive' interpretation that 'covers' all of the alternative interpretations, similar to those proposed for certain cases of lexical polysemy in (Copestake and Briscoe, 1995). We will hypothesize below that the existence of such an underspecified interpretation may be a further important property of these contexts.

The third property that these examples have in common is that both in situations involving attaching two objects together and in situations involving loading objects into other objects, all of the alternative interpretations of the anaphoric expression are equivalent as far as the plan (moving these objects to a new location) is concerned: after the two explicitly mentioned potential antecedents are joined, if one of them gets moved, the other one must be moved as well. E.g., in (7), 27.4, all interpretations of the instruction *send that off* will achieve the same result irrespective of how the pronoun is interpreted. We will write X $\sim$ Y to indicate that interpretation X is equivalent to interpretation Y for the purpose of the plan: i.e., we write

$$oj \sim tc$$

to indicate that from the point of view of the plan, interpreting the pronoun *that* as referring to the orange juice or the tanker car are equivalent. Similarly, in the case of (5), we will write $e \sim b$ to say that from the point of view of the plan, the interpretation of *it* in which it refers to engine E2 and that in which it refers to the boxcar are equivalent.

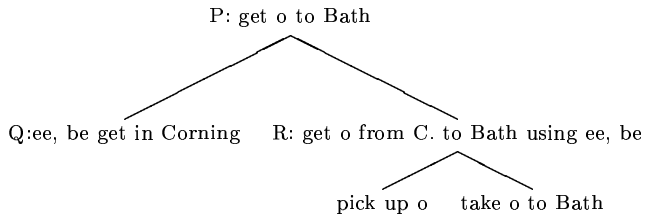
3.3. Reference to plans

A second class of pronominal uses in the TRAINS dialogues satisfies our three conditions (ambiguity, no preferences for one interpretation, and no complaints). These are the uses of demonstratives such as *that* to refer to parts of a plan, as in the following example (from TRAINS dialogue d91-1.1):

(8)
```
11.6     : aha
11.7     : I see an engine and a boxcar both
           at Elmira
12.1    S: right
13.1    M: this looks like the best thing
           to do
13.2     : so we should get
13.3     : ... the eng / engine to picks up
           the boxcar
13.4     : and head for Corning
13.5     : 's that sound reasonable
14.1    S: sure
14.2     : that sounds good
15.1    M: and from Corning we'll pick up
           the oranges
15.2     : and um
15.3     : take them to Bath
15.4     : will it / that get m / me
15.5     : do you think that I can get  ..
           this all over to Bath by 8 o'clock
16.1    S: yeah
16.2     : that gets us to Bath at f / 5 AM
16.3     : so it's plenty of time
```

The demonstratives in question are *that* in 15.4 and *that* in 16.2. Roughly speaking, the structure of the plan at this point in the conversation can be represented as follows:

(9)



but from the transcript it is unclear whether these two demonstratives refer just to the portion of the plan consisting of the action of picking up the oranges and taking them to Bath (or perhaps just to one of the two actions, say the second), or to the larger plan which also includes the previous action of getting the engine to pick up the boxcar and heading for Corning. Our two other conditions are met as well: no focusing principle we are aware of suggests one interpretation over the other, and the other participant does not complain.

These examples do not look like cases of vagueness, either, as shown by the fact that (10) does not have the interpretation that John could agree to a plan, whereas Bill would agree to a subpart of it:

(10) *John agree to THAT PLAN/THAT, and Bill did too*;

The diagrammatic representation of the plan in (9) already hints at a further semantic property that plans share with the antecedents of the previously discussed class of pronominal use: plans, at least as classically viewed in the Artificial Inteligence literature, have a mereological structure as well. According to this view of plans, plans denote event or action types, and these in turn can be seen as sets of actions (of the appropriate type). So, for example, plan R in (9) denotes the set of events of getting oranges from Corning to Bath using the engine and the boxcar,

$$\mathrm{R} = \{e|e{:}\mathbf{get}(\mathrm{o,c,ba,ee,be})\}$$

where *o* refers to the oranges, *c* referes to Corning, *ba* refers to Bath, *ee* refers to the engine at Elmira, and *be* to the boxcar at Elmira. Plan P refers to the set of events of getting oranges to Bath:

$$\mathrm{P} = \{e'|\exists x, y, z e'{:}\mathbf{get}(\mathrm{o,x,ba,y,z})\}$$

Intuitively, set R is a subset of set P. This allows us to define a 'part-of' relation $\lhd$ between plans, to be interpreted as PLAN DECOMPOSITION, as follows. Let $\subseteq$ be the relation of inclusion between events; then R is a part of P iff for every event e in R there is an event e' in P such that $e \subseteq e$, as follows:

$$R \lhd P \equiv [\forall e \in R, \exists e' \in Pe \subseteq e']$$

With this interpretation of plans, we can see that the potential antecedents of the ambiguous demonstratives are part of a mereological structure similar to that observed in the previous examples, and that there is a similar range of possibilities concerning the interpretation derived by the listener. In particular, a p-underspecified interpretation is available, as in the previous cases. One crucial difference is that in the case of reference to plans one interpretation does not seem available: this is the interpretation in which the pronoun refers to Q. This appears to be yet another instance of the so-called 'right frontier constraint' often discussed in the literature on references to abstract objects (Webber, 1991). As a result, the p-underspecified interpretation now appears to be further constrained, as well: the antecedent z of the pronoun is not merely dominated by the supremum of the current plan, P; but it's also part of its right frontier:

$$z \lhd^* P,\ z \in RF(P)$$

What's more, these last examples have a further, and crucial, similarity to the mereology cases discussed earlier: in these cases, as well, we can say that the two possible interpretations of the relevant utterance (e.g., 15.4, *that gets us to Bath at 5 AM*) are equivalent for the purpose of the plan. This is because once the part of the plan being proposed by M in 13.1-13.5, Q, has been accepted by S (utterances 14.1-14.2), whether or not the plan as a whole (P) is going to work depends entirely on whether subplan R is going to work; so accepting R is essentially equivalent to accepting P as a whole:

$$P \sim R$$

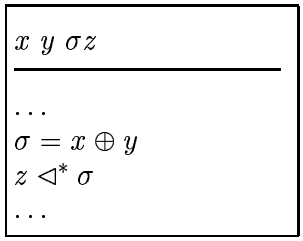
4. The Justified Sloppiness Hypothesis

The discussion in the previous section makes it clear that the two patterns of anaphoric reference we have observed in the TRAINS dialogues have at least three aspects in common:

1. Both explicitly mentioned potential antecedents x and y are elements of an underlying mereological structure with summum $\sigma = x \oplus y$ which has been explicitly constructed (and made salient) in the dialogue ($\sigma = oj \oplus tc$ in (7), $\sigma = P$ in the case of (8));

2. the existence of this structure makes it possible to construct a p-underspecified interpretation in which the anaphoric expression is

interpreted as denoting an element z included in the mereological structure - i.e., part-of its summum σ:



3. All possible interpretations $(x, y, z, x \oplus y)$ are equivalent for the purposes of the plan.

This suggest the following preliminary hypothesis:

> Ambiguous anaphoric expressions are not perceived as infelicitous provided that Conditions 1-3 hold.

This may be because if these three conditions hold, the speaker's sloppiness in using an anaphoric expression in an ambiguous context is not problematic; we will therefore use the term JUSTIFIED SLOPPINESS to indicate cases such as those discussed in the previous section, and refer to the hypothesis above as **Justified Sloppiness Hypothesis**, or JSH.

Of course, the fact that a p-underspecified interpretation exists does not mean that the listener will adopt it as its final interpretation; however, this possibility is what makes these examples interesting from an underspecification perspective.

5. Additional cases of justified sloppiness

5.1. EVENTS RELATED BY GENERATION

After identifying the cases discussed above, we discovered that similar examples of potentially ambiguous anaphoric expressions which, however, did not appear to be problematic for the reader had already been discussed by Schuster (1988), albeit not in connection with the question of whether anaphoric reference is underspecified or not.[2] Schuster analyzes two types of data: dialogues between an expert and a novice attempting to learn how to use the Emacs editor, and questionnaires asking subjects to indicate the preferred referents of anaphoric expressions referring to events. The transcripts of Emacs dialogues include several examples of reference to events, the most interesting

[2] These examples were pointed out to us by Bonnie Webber.

among which for our purposes are examples like the following (our own indices):

(11) a. E: Do this: [$_1$ set a "mark" at some point (any old point) by [$_2$ typing <esc>-M]]. It will say "mark set". Try it.

b. E: <esc>-M will give set-mark. Try it.

Schuster uses such examples to argue for a representation of events in which an event may GENERATE another event, such as those developed by Goldman (1970) and Pollack (1986). Examples like those in (11) are used as diagnostics, in the sense that the possibility of using a single pronoun indicates that the two events are related by a generation relation:

> In both cases [our (11a) and (11b), NDR] the referent(s) of the pronoun *it* can be either "setting the mark" or "typing <esc>-M" or even both: "setting the mark by typing <esc>-M". ... we can claim that "typing <esc>-M at a given time" can generate "setting the mark at that given time" ... ((Schuster, 1988), p.9-10).

Schuster also observes that these references appear to be unproblematic; her explanation is:

> This relationship [generate] allows us to establish a connection between "typing <esc>-M" and "setting the mark"and it can be understood as one relationship ... When the pronoun *it* is used as is the case in both examples, neither of the two referents need to be specified because the generation relationship indicates that they are both related to each other ...

She also argues that if the generation relation is not properly established, such references may turn out to be ambiguous, as in the following example:

(12) Set the mark at the beginning of the region. Type <esc>-M and once you've done $that_1$, move to the end of the region.

These cases bear a considerable resemblance to those discussed in the previous section: two actions that are closely 'tied together' and as a result reference to the one becomes equivalent to a reference to the other. Using *tem* to indicate "typing <esc>-M"and *stm* for setting the mark, we can say again that the two interpretations for the discourse entity z for *that* are equivalent:

$$tem \sim stm$$

The difference in this case is that instead of a 'part-of' relationships between the actions, as in the case of reference to plans discussed

earlier, we have a much tighter relationship: generation makes the two actions almost into a single action. As a result, in this case we cannot even talk of 'sloppy' references: these references are perfectly accurate. We can still assume that the pronoun gets assigned an underspecified interpretation as discussed above, but this interpretation is almost not 'underspecified':

$$z \lhd^* (stm \oplus tem), \textbf{generates}(stm,tem)$$

5.2. An unclear case: References to spatial areas

We are also aware of a few cases that cannot clearly be reconduced under our generalization, either because it isn't clear whether the reference was truly ambiguous or merely vague, or because of lack of evidence concerning whether readers truly find these references unproblematic. Poesio and Vieira (1998) report that human subjects do not agree on the interpretation of definite descriptions such as *the area* in the following example:

(13) *About 160 workers at a factory that made paper for the Kent filters were exposed to asbestos in the 1950s. Areas of the factory were particularly dusty where the crocidolite was used. Workers dumped large burlap sacks of the imported material into a huge bin, poured in cotton and acetate fibers and mechanically mixed the dry fibers in a process used to make filters. Workers described "clouds of blue dust" that hung over parts of the factory, even though exhaust fans ventilated the area.*

Three subjects were asked to indicate the antecedent of this description in the text. One subject indicated *parts of the factory* as the antecedent; another indicated *the factory*; and the third indicated *areas of the factory*.

In this example, again, we have an underlying mereological structure: both *parts of the factory* and *areas of the factory* are obviously included in the total area of the factory. There are, however, a few problematic issues in these examples. First of all, in this case there is no obvious equivalence between the three different interpretations. Furthermore, in this case one could argue that the object referred to - the area - is not in focus; to the extent that one can say that there is a focus in this text, it is most likely the factory. So, given the results of the experiments of Garrod *et al.*, this is perhaps the example in which it is most likely that the reader did not even attempt to construct an interpretation for the anaphoric expression.

6. The Limits of Corpus Analysis

It is a good idea to stop at this point and think again about which questions we would like to see answered, and what we can expect from an analysis of corpus data like the one we just presented. With corpus analysis, we can identify contexts in which 'sloppy references' take place, and commonalities between the semantic interpretations produced in each case. But it is important to realize that most of the important questions cannot be answered this way.

In all of the examples we discussed the speaker is taking a risk, yet the listener does not signal a problem in understanding. An interesting question raised by this observation is whether the speaker is simply being sloppy, or he/she has done what Hobbs (1985) would call 'collapsing a complex theory in one of coarser granularity'–i.e., he/she is aware that there are two possible interpretations, but is also aware that the two interpretations are equivalent. This is one example of a question that cannot be answered by corpus analysis –or indeed, by any other technique we are aware of, because doing so would require reading the mind of the speaker.

More amenable questions are whether it's really the case that listeners do not find these cases problematic, and if so, what kind of interpretations they are constructing. The space of possible answers to this second question is as follows:

1. The listener doesn't even attempt to interpret the pronoun, and keeps what is called a H-UNDERSPECIFIED INTERPRETATION in (Poesio, 1999): i.e., an interpretation in which the conventional meaning of some sub-utterances has not been determined and, as a consequence, the conventional meaning of the utterance as a whole is not determined. This hypothesis would be perhaps more plausible in the case of non-task-oriented dialogues such as those in the Switchboard corpus; less so in the case of task-oriented dialogues. Furthermore, the results by Garrod et al. (1994) suggest that listeners do interpret pronouns when the entity they refer to is in focus (as is the case in all examples here).[3]

2. The listener does attempt to interpret the pronoun. Again, there are two possibilities:

   a) The listener realizes that there are two possible interpretations.

[3] Perhaps one could argue that at this point in the conversation the listener (usually S) is simply constructing a very rough plan, without really trying to interpret everything that the speaker says; this is left for later. This hypothesis is, however, hard to distinguish from Hypothesis 2.a.

In this case, there are three more possibilities:

i) The listener realizes that the two objects are part of a same mereological structure t; so it builds a (P-UNDERSPECIFIED) interpretation (Poesio, 1999) where the pronoun is assigned a discourse entity z as interpretation, with the constraints that
$z \lhd (e \oplus b)$, $ATOM(z)$.

ii) The listener performs a shift in granularity, building a new interpretation in which e and b are treated as the same object.

iii) The listener interprets the pronoun as referring to the mereological structure itself, $(e \oplus b)$.

iv) The listener chooses one of the two interpretations, whether or not he/she realizes that they are equivalent (he/she may also be taking a risky strategy and 'hope for the best').

b) The listener only finds one possible interpretation for the pronoun, either e or b; no communication problem ensues, since the two interpretations are equivalent.

Corpus analysis can't answer these questions, either; but in this case, we are talking about questions that may be answerable using controlled psychological experiments. In the next section, we discuss our preliminary experimental results.

7. Testing the Justified Sloppiness Hypothesis

The Justified Sloppiness Hypothesis is a fairly weak claim, in that it does not say anything concerning the actual interpretation of the pronominal uses we identified: it merely asserts that our 'lack of problem signals' heuristic is correct, and that cases of pronominal reference that satisfy the conditions we identified are indeed felicitous. As such, it is fairly easy to check: we simply have to test whether sentences that contain a potentially ambiguous anaphoric reference are easier to process when the two potential antecedents are part of a mereological structure rather than being separate. A number of techniques can be used to test hypotheses of this kind; we used the Magnitude Estimation technique proposed in (Bard et al., 1996).

Methods
To test the JSH, we asked subjects to judge whether sentences such as (14a) are 'more acceptable' (in that less ambiguous) than the minimally

different (14b), in which *the engine* and *the boxcar* are not attached together:

(14) a. The engineer hooked up the engine to the boxcar and sent it to London.

b. The engineer separated the engine from the boxcar and sent it to London.

In Magnitude Estimation experiments, the subjects are asked to assign a magnitude (an arbitrary number) to a reference sentence, and then have to judge the acceptability of other sentences relative to the reference magnitude.

This experiment involves two conditions: 'MEREOLOGY' (M) and 'NON-MEREOLOGY' (NM). To compare these two conditions we used minimal pairs of the form shown in (14), with an identical second part containing the anaphoric expression, and first parts that differ only in the verb: in the MEREOLOGY condition, a verb suggesting that the two objects are part of a larger block is used (e.g., *hooked up .. to*; in the NON-MEREOLOGY condition, that the two objects are disjoint (e.g., *separated ... from*). We adopted a Latin Square design, whereby each subject sees only one element of the minimal pair. (In fact, we also asked our subjects to estimate the acceptability of the first parts only, to make sure the differences were not there. As a result, we got four groups of subjects.)

The experiment was run using WebExp, a software package for running experiments on the Web developed at the Universities of Edinburgh and Saarbruecken (`http://www.hcrc.ed.ac.uk/web_exp/`). The subjects connect to a web page (`http://www.cogsci.ed.ac.uk/~poesio/web_exp/undersp1.instr.html`) and follows the instructions at her/his own pace. We had 28 subjects in total.

Results

We found a significant effect of mereology both on a by-subject and a by-item analysis.[4] The means for the by-subject analysis are shown in Table 7.

A two-way ANOVA test over these means indicates a Length effect (First Parts are more acceptable than Full Materials) of no concern to us, but also that Mereology items like (14a) are significantly more acceptable than Non-Mereology items like (14b): $F_s(1,27) = 36.78$ ($p < 0.000$). Crucially, we only find this effect when comparing Full Ma-

[4] A by subjects analysis indicates whether the results generalize across subjects– i.e., whether new subjects are likely to behave like the ones we tested. A by items analysis indicates whether the results generalize across materials.

Table I. By-subject means for Experiment 1

	Mereology	Nonmereology	Total
Full Material	0.0409	-0.0656	-0.0123
First Part	0.1712	0.1403	0.1557
Total	0.1061	0.0373	

terials, not when comparing First Parts: $F_s = 7.45$ ($p < 0.011$) for interaction M x P. Similar results are obtained when analyzing the means by items: $F_a = 9.43$ ($p < 0.005$) for Mereology, $F_a = 5.196$ ($p < 0.032$) for interaction M x P.

Discussion
These results support the JSH; preliminary analyses of these results also suggest that a few refinements may be necessary. First of all, we observed that the availability of the 'underspecified' interpretation discussed above is affected by the salience of the antecedents: when the two antecedents are highly salient, the interpretation becomes difficult, and the sentences less acceptable. For example, entities introduced by proper names are 'too salient':

(15) Sue tied John's bike to Bill's bike. It wouldn't move anymore.

Also, there seems to be a difference between instructions and assertions: in the first case, there seems to be a stronger feeling that a plan is being developed, which seem to make condition 2 easier to achieve.

The next step is to address the second–and central– question: what kind of interpretation is being produced by the listener in these cases? This is the goal of a follow-up experiment currently under way, also using Web Exp, which involves the same materials, but in which we ask our subjects to indicate the antecedents by means of a multiple-choice questionnaire.

8. Conclusions

In summary, we analyzed a corpus of task-oriented dialogues, finding two patterns of anaphoric pronoun usage in which apparently ambiguous pronouns do not seem to result in communication problems: references to objects which are part of a larger mereological structure, and references to events which are part of a plan. Our analysis of

these cases led us to formulate a *Justified Sloppiness Hypothesis*, stating that such apparently ambiguous uses are felicitous provided that the following three conditions hold:

1. The alternative interpretations are part of an underlying mereological structure, which has been made salient in the discourse;
2. this structure allows a p-underspecified interpretation;
3. the alternative interpretations are equivalent for the purposes of the plan.

We also argued that examples previously noted by Schuster can be reconduced under the proposed generalization. Controlled psychological experiments did support the JSH, showing that pronominal uses satisfying these three conditions are significantly more acceptable than analogous ambiguous cases in which no mereological structure is made salient.

This is, of course, only a first step. Our future research plans include testing whether the p-underspecified interpretations allowed by the discourse model are actually chosen as the interpretation of the pronouns–a difficult question to answer without forcing our subjects to choose one of these interpretations. It also turns out that speakers are not always so careful; there also seem to be UNJUSTIFIED SLOPPINESS cases, in which only the first Condition in the JSH is satisfied, yet speakers still use pronouns. These cases, as well, we plan to study in the future.

Acknowledgments

Thanks to Ellen Bard, Antje Roßdeutscher, Hannes Rieser, Patrick Sturt, and Bonnie Webber for comments and suggestions; special thanks to Frank Keller and Patrick Sturt for help with the design of the experiments. This work was supported in part by Advanced Research Fellowship B/96/AF/2266 from the UK Engineering and Physical Sciences Research Council (Poesio), in part by a European Science Exchange Programme grant from the Royal Society, *Cases of Unresolved Underspecification* (Poesio and Reyle).

References

Alshawi, H. and R. Crouch: 1992, 'Monotonic Semantic Interpretation'. In: *Proc. 30th. ACL.* University of Delaware, pp. 32–39.

Bard, E. G., D. Robertson, and A. Sorace: 1996, 'Magnitude Estimation of Linguistic Acceptability'. *Language* **72**(1), 32–68.

Copestake, A. and T. Briscoe: 1995, 'Semi-Productive Polysemy and Sense Extension'. *Journal of Semantics* **12**(1), 15–68. Special Issue on Lexical Semantics.

Frazier, L. and K. Rayner: 1990, 'Taking on Semantic Commitments: Processing Multiple Meanings vs. Multiple Senses'. *Journal of Memory and Language* **29**, 181–200.

Garrod, S. C., D. Freudenthal, and E. Boyle: 1994, 'The role of different types of anaphor in the on-line resolution of sentences in a discourse'. *Journal of Memory and Language* **32**, 1–30.

Garrod, S. C. and A. J. Sanford: 1985, 'On the real-time character of interpretation during reading'. *Language and Cognitive Processes* **1**, 43–61.

Gernsbacher, M. A. and D. Hargreaves: 1988, 'Accessing Sentence Participants: The Advantage of First Mention'. *Journal of Memory and Language* **27**, 699–717.

Goldman, A.: 1970, *A Theory of Human Action.* Princeton, NJ: Princeton University Press.

Grosz, B. J., A. K. Joshi, and S. Weinstein: 1995, 'Centering: A Framework for Modeling the Local Coherence of Discourse'. *Computational Linguistics* **21**(2), 202–225. (The paper originally appeared as an unpublished manuscript in 1986.).

Hobbs, J. R.: 1985, 'Granularity'. In: *Proceedings of the Ninth International Joint Conference on Artificial Intelligence.* Los Angeles, California, pp. 432–435.

Lakoff, G. P.: 1970, 'A note on vagueness and ambiguity'. *Linguistic Inquiry* **1**(3), 357–359.

Link, G.: 1983, 'The Logical Analysis of Plurals and Mass Terms: A Lattice- Theoretical Approach'. In: R. Bäuerle, C. Schwarze, and A. von Stechow (eds.): *Meaning, Use and Interpretation of Language.* Walter de Gruyter, pp. 302–323.

Muskens, R.: 1995, 'Order-independence and underspecification'. In DYANA-2 Deliverable R2.2.C, *Ellipsis, Underspecification, and Events in Dynamic Semantics.*

Pinkal, M.: 1995, 'Radical Underspecification'. In: P. Dekker, J. Groenendijk, and M. Stokhof (eds.): *Proceedings of the Tenth Amsterdam Colloquium.*

Poesio, M.: 1991, 'Relational Semantics and Scope Ambiguity'. In: J. Barwise, J. M. Gawron, G. Plotkin, and S. Tutiya (eds.): *Situation Semantics and its Applications, vol.2.* Stanford, CA: CSLI, Chapt. 20, pp. 469–497.

Poesio, M.: 1994, 'Discourse Interpretation and the Scope of Operators'. Ph.D. thesis, University of Rochester, Department of Computer Science, Rochester, NY.

Poesio, M.: 1996, 'Semantic Ambiguity and Perceived Ambiguity'. In: K. van Deemter and S. Peters (eds.): *Semantic Ambiguity and Underspecification.* Stanford, CA: CSLI, Chapt. 8, pp. 159–201.

Poesio, M.: 1999, 'Utterance Processing and Semantic Underspecification'. HCRC/RP 103, University of Edinburgh, HCRC.

Poesio, M.: 2004, *Incrementality and Underspecification in Semantic Interpretation*, Lecture Notes. Stanford, CA: CSLI. To appear.

Poesio, M. and R. Stevenson: To appear, *Salience: Theoretical Models and Empirical Evidence.* Cambridge and New York: Cambridge University Press.

Poesio, M. and R. Vieira: 1998, 'A Corpus-Based Investigation of Definite Description Use'. *Computational Linguistics* **24**(2), 183–216. Also available as Research Paper CCS-RP-71, Centre for Cognitive Science, University of Edinburgh.

Pollack, M. E.: 1986, 'Inferring Domain Plans in Question-Answering'. Ph.D. thesis, Department of Computer and Information Science, University of Pennsylvania.

Reyle, U.: 1993, 'Dealing with ambiguities by underspecification: Construction, Representation and Deduction'. *Journal of Semantics* **10**, 123–179.

Reyle, U.: 1996, 'Co-indexing Labeled DRSs to Represent and Reason with Ambiguities'. In: K. van Deemter and S. Peters (eds.): *Semantic Ambiguity and Underspecification*. Stanford: CSLI, Chapt. 10, pp. 239–268.

Sanford, A. J. and P. Sturt: 2002, 'Depth of processing in language comprehension: not noticing the evidence'. *Trends in Cognitive Science* **6**, 382–386.

Schuster, E.: 1988, 'Pronominal reference to events and actions: Evidence from naturally-occurring data'. LINC LAB 100, University of Pennsylvania, Dept. of Computer and Information Science, Philadelphia.

Sidner, C. L.: 1979, 'Towards a computational theory of definite anaphora comprehension in English discourse'. Ph.D. thesis, MIT.

Stevenson, R. J., R. A. Crawley, and D. Kleinman: 1994, 'Thematic Roles, Focus, and the Representation of Events'. *Language and Cognitive Processes* **9**, 519–548.

Sturt, P. and M. Crocker: 1996, 'Monotonic Syntactic Processing: A cross-linguistic study of attachment and reanalysis'. *Language and Cognitive Processes* **11**(5), 449–494.

Swinney, D. A.: 1979, 'Lexical Access During Sentence Comprehension: (Re)consideration of Context Effects'. *Journal of Verbal Learning and Verbal Behavior* **18**, 545–567.

Tanenhaus, M. K., M. Spivey-Knowlton, K. M. Eberhard, and J. C. Sedivy: 1995, 'Integration of Visual and Linguistic Information in Spoken Language Comprehension'. *Science* **268**, 1632–1634.

van Deemter, K. and S. Peters (eds.): 1996, *Semantic Ambiguity and Underspecification*. Stanford: CSLI Publications.

van Eijck, J. and J. Jaspars: 1996, 'Underspecification and Reasoning'. In *Building the Framework*, Deliverable D15 of the FRACAS project. Available at URL `http://www.cogsci.ed.ac.uk/~fracas/`.

Webber, B. L.: 1991, 'Structure and Ostension in the Interpretation of Discourse Deixis'. *Language and Cognitive Processes* **6**(2), 107–135.

Zwicky, A. and J. Sadock: 1975, 'Ambiguity Tests and How to Fail Them'. In: J. Kimball (ed.): *Syntax and Semantics 4*. New York: Academic Press, pp. 1–36.