Dialogue and Discourse 2(1) (2011) 235–277 doi: 10.5087/dad.2011.110

An Incremental Model of Anaphora and Reference Resolution Based on Resource Situations

Massimo Poesio
University of Essex and Università di Trento

Hannes Rieser
Universität Bielefeld

Editor: David Schlangen

Abstract

Notwithstanding conclusive psychological and corpus evidence that at least some aspects of anaphoric and referential interpretation take place incrementally, and the existence of some computational models of incremental reference resolution, many aspects of the linguistics of incremental reference interpretation still have to be better understood. We propose a model of incremental reference interpretation based on Loebner's theory of definiteness and on the theory of anaphoric accessibility via resource situations developed in Situation Semantics, and show how this model can account for a variety of psychological results about incremental reference interpretation.

1. Introduction

Evidence from both corpora and behavioral experiments suggests that at least some aspects of the interpretation of referring expressions are incremental. For instance, in the following fragment from the TRAINS corpus of dialogues collected at the University of Rochester by the TRAINS project (Allen et al. 1995),[1] the repair in utterance 10.1 is clearly initiated because participant S has started processing the definite description *the engine at Avon* before M's utterance is complete, and has identified the actual referent of the definite description (engine E1).

(1)
```
9.1  M: so we should
9.2   : move the engine
9.3   : at Avon
9.4   : engine E
9.5   : to
10.1 S: engine E1
11.1 M: E1
12.1 S: okay
13.1 M: engine E1
13.2  : to Bath
13.3  : to /
13.4  : or
13.5  : we could actually move it
```

1. The methodology employed in compiling the transcripts, including the methodology used for segmenting utterances into utterance units approximately expressing prosodic boundaries, is described in (Gross et al. 1993).

 Submitted 1/2010; Accepted 3/2011; Published online 5/2011

```
          to Dansville to pick up
          the boxcar there
    14.1 S: okay
```

Substantial behavioral evidence from the last fifteen years conclusively supports the intuitions gained by studying such corpus data. In particular, studies using the visual world paradigm have shown that subjects following spoken instructions to manipulate objects in a "visual world" fixate to the relevant objects as soon as the phonetic prefix is completely unambiguous (Tanenhaus et al. 1995; Eberhard et al. 1995; Tanenhaus and Trueswell 2005). For example, in a visual world containing an apple and a towel, subjects will fixate on the towel as soon as they hear the first syllable of the word *towel*.

Although accounts of incremental interpretation of referring expressions have been developed—e.g., (Stoness et al. 2004; Schlangen et al. 2009; Dubey 2010)—the relation between psychological evidence and existing linguistic theories of anaphora and reference is still poorly understood. In this paper, we propose a theory of the incremental interpretation of referring expressions in terms of a theory of the linguistics of such expressions based, on the one hand, on the theory of definiteness developed by Loebner (1987); on the other, on the theory of anaphoric accessibility via resource situations developed in Situation Semantics (Barwise and Perry 1983; Gawron and Peters 1990; Poesio 1993; Cooper 1996). The paper builds on previous work in the PTT framework (Poesio and Traum 1997; Poesio and Rieser 2010), but makes two novel contributions. First, it consolidates into a single proposal a number of ideas about definites, resource situations, and anaphora developed over the years within PTT but never integrated into a coherent whole. Secondly, it provides an explicit account of the main psychological evidence about reference interpretation gathered using the visual world paradigm.

The structure of the paper is as follows. In section 2 we summarize the main psychological evidence about incrementality and reference. In section 3 we provide a quick summary of the aspects of PTT from (Poesio and Rieser 2010; Poesio To Appear) essential for the present proposal. In section 4 we provide a novel and unified account of the semantics and pragmatics of referring expressions. Finally, in section 5 we show how the proposal can explain the evidence discussed in section 2. A survey of related literature and a discussion follow.

2. Psychological Evidence on Incrementality and Reference

In this section we discuss the key evidence about incremental interpretation in general, and in particular the evidence about incremental reference resolution that our theoretical proposals were designed to explain.

2.1 Combinatorial Explosion and Incrementality

Given the number of phonetically, lexically, syntactically, semantically and pragmatically distinct readings identified by theoretical linguists for most natural language expressions, it is a wonder that such expressions can be understood at all. Yet, people appear able to process them rapidly and without apparent effort. The explanation for this puzzle involves a variety of factors, but clearly, one of the key ingredients of the solution is the fact that people appear to process natural language expressions incrementally, immediately making choices about their interpretation before proceeding to the next input segment.

The initial evidence about the incremental nature of human language processing came from research on human parsing, in the form of the phenomenon of **garden paths** observed by Bever (1970). Garden paths are sentences such as those in (2), which are perfectly grammatical, but which subjects nevertheless find odd because the ambiguity between a reduced relative reading and a matrix verb reading of the verbs *raced*, *floated* etc. is immediately resolved in favor of the matrix verb interpretation, thus forcing the reader to a reanalysis step later on.

(2) a. *The horse raced past the barn fell.*

b. *The boat floated down the river sank.*

Shortly after the initial findings by Bever, psychological evidence was found suggesting that semantic interpretation processes are incremental, as well. Well-known cross-modal priming experiments suggested that lexical access proceeds by first immediately activating all senses of an ambiguous word-form, and then immediately discarding all but the chosen one (Swinney 1979; Seidenberg et al. 1982). In these experiments, the subjects were presented with texts such as the one in (3); half of the time a disambiguating context was provided (the string *spiders, roaches and other*). Swinney found priming effects for both *ant* and *spy* at [1], even with a strongly disambiguated context; but only for *ant* at [2].

(3) *Rumour had it that for years the government building had been plagued with problems. The man was not surprised when he found several (spiders, roaches, and other) bugs [1] in the corner [2] of his room.*

Using similar methods, Corbett and Chang (1983) found that anaphora resolution, as well, involved the rapid activation of all (gender- and number-) matching antecedents of anaphors like *she* in (4a), as could be verified by cross-modal testing of the activation of the antecedents at [1]. All but one of these however were dropped by point [2] in different-gender conditions, but not in same-gender conditions as in (4b).

(4) a. Karen poured a drink for Bob and then Karen, she [1] put [2] the bottle down.

b. Karen poured a drink for Emily and then Karen, she [1] put [2] the bottle down.

2.2 Parallelism

The original data from Bever led to the development of so-called **garden-path theory** (Frazier 1979, 1987) and numerous other incremental models of parsing based on the assumption that interpretations were generated in a serial fashion, one at a time (Abney 1991; Shieber and Johnson 1993; Milward 1994). However, the cited results about lexical access could only be explained in terms of parallel processing (Marslen-Wilson 1975), and indeed they led to the development of the so-called **cohort** model of lexical access (Marslen-Wilson 1987). The results by Corbett and Chang about pronominal interpretation, as well, suggest a parallel model. In recent years, the predominant view has been that parallel processing is the rule in the case of syntax as well (Gibson 1991; Jurafsky 1996; Pearlmutter and Mendelsohn 1999); recent evidence on the relation between parsing and lexical disambiguation also suggests a parallel model (MacDonald et al. 1994).

2.3 Incrementality in Reference: The Visual World Paradigm

As discussed above, early evidence that anaphora resolution is incremental was provided by the cross-modal priming experiments by Corbett and Chang (1983). These results were confirmed and much strengthened by work using the so-called **visual world paradigm** (Tanenhaus et al. 1995; Eberhard et al. 1995; Arnold et al. 2000; Tanenhaus and Trueswell 2005).

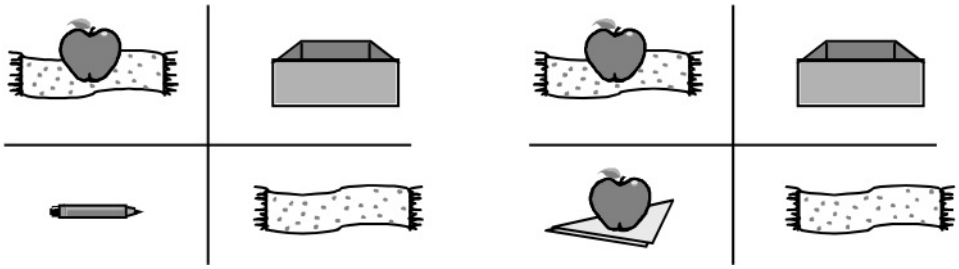


Figure 1: The Visual World Paradigm: Ambiguous and unambiguous visual world situations

In the visual world paradigm, subjects looking at scenes such as those in Figure 1 (this is Figure 4 from (Tanenhaus et al. 2004)) and wearing head-mounted eye trackers hear instructions such as (5). Through the eye trackers it is possible to measure the percentage of fixations to each object in the visual scene millisecond by millisecond; a concentration of fixations on a given object from a certain point on (typically around 400ms after the onset of the target referring expressions) provides very good evidence that that object has been identified as the referent of the expression.

(5) Put the apple on the towel in the box.

Eberhard et al. (1995) and Allopenna et al. (1998) showed that this concentration of fixations on the referent occurs as soon as an unambiguous prefix has been processed—i.e., in a visual scene in which there is only one object whose name begins with *ap-*, fixations begin to concentrate on that object as soon as that prefix has been processed, without waiting for the rest of the head noun. In fact, Eberhard et al. (1995) showed that in cases in which the referring expression contains unambiguous modifiers, subjects do not even wait until the head noun before beginning to concentrate their attention—i.e., in a situation in which there is only one red object, fixations concentrate on that object 400ms after the onset of *red*.

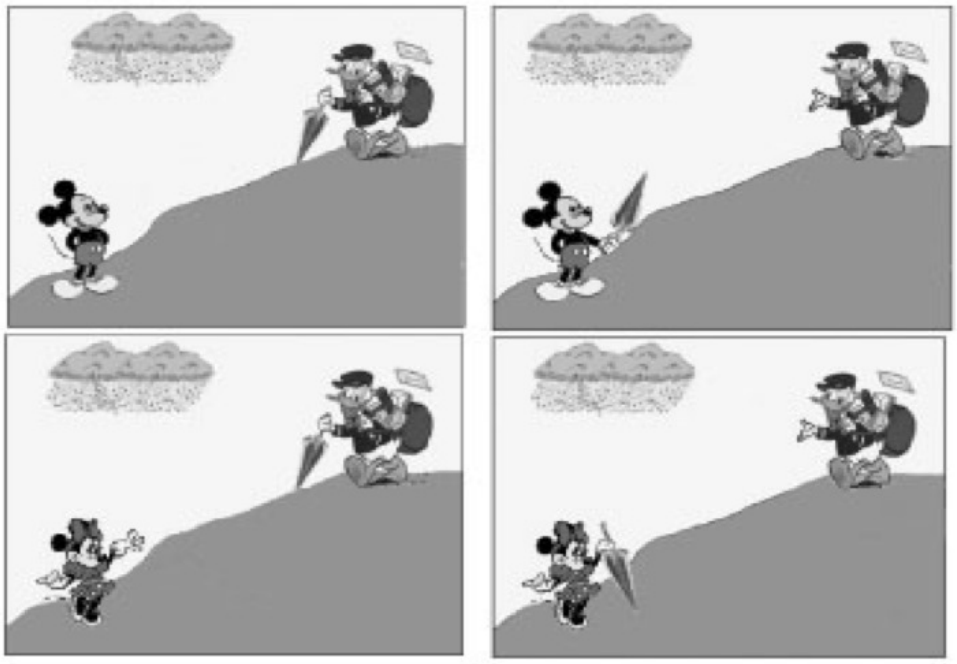


Figure 2: The materials used by Arnold *et al.* to study pronouns with the Visual World Paradigm

The first studies of reference using the visual world paradigm focused on the interpretation of nominals. Arnold et al. (2000) extended the use of the paradigm to the study of the interpretation of pronouns. The subjects of Arnold *et al.* listened to two-sentence texts while viewing one of the four pictures in Figure 2. The first sentence of the text contained either two same-gender referents (Donald / Mickey) or two different-gender ones (Donald / Minnie), where the second sentence contained either a masculine or a feminine pronoun, as in (29).

(6) a. Donald is bringing some mail to [Mickey / Minnie]
while a violent storm is beginning.

b. He's / she's carrying an umbrella,
and it looks like they're both going to need it.

Arnold *et al.* found both a gender and a first-mention effect. In the different gender contexts, fixations would concentrate on the unambiguous referent of the pronoun already after 400ms, and so they would in the same gender context when reference was to the first mention entity. In the same gender, second mention reference context, however, the percentage of fixations on the first and second mentioned entity was the same.

Finally, a series of visual world paradigm experiments including, among others, (Altmann and Kamide 1999; Chambers et al. 2002; Brown-Schmidt et al. 2005) found substantial empirical evidence for the **focus shift principles** proposed in (Grosz 1977; Poesio 1993) on the basis of the analysis of data from task-oriented dialogues, and later studied by Beun and Cremers (1998). These

focus shifting effects have now become known as effects of **referential domain restriction** (Brown-Schmidt et al. 2005).

Chambers et al. (2002) found that after hearing instruction (7) in a visual scene containing a number of containers some of which are big enough to fit the cube whereas others aren't, the subjects' attention quickly concentrates on the containers into which the cube can fit. This type of referential domain restriction / focus shifting effect is now known as **task compatibility**.

(7) Pick up the cube. Put it in ...

The experiments by Chambers *et al.* took place in controlled experimental situations. The experiments discussed in (Brown-Schmidt et al. 2005; Brown-Schmidt and Tanenhaus 2008), by contrast, involved subjects performing tasks in fairly ecologically valid situations; but these studies, as well, found focusing effects—in particular, effects both of task compatibility in the sense of Chambers *et al.* and proximity (greater salience of closest objects).

2.4 Interaction between Reference and Parsing

Crain and Steedman (1985) and Altmann and Steedman (1988) observed that many classical garden path sentences such as (2) or (5) involve an ambiguity between a reading in which a constituent (*raced past the barn*, *on the towel*) is interpreted as a modifier of a definite NP and a second reading in which it is interpreted as part of the main clause. They also observed that the fact that this second reading is initially preferred—thus originating the garden path—might be due to the lack of a second object in the context (a second horse, or a second apple) that would justify the use of the modification; and hypothesized that the garden path effect might be reduced, or eliminated, in contexts in which this object is present.

The visual world paradigm offered the opportunity for a very convincing verification of this hypothesis, reported in Tanenhaus et al. (1995) and Spivey et al. (2002). The subjects in these experiments were presented either with the visual context on the left in Figure 1, in which only one apple is present, or with the context on the right, in which there are two. They then heard the instruction in (5) on page 230. A much greater proportion of fixations on the incorrect destination (the towel) was observed in the situation in which only one apple was present.

3. A Short Introduction to PTT

PTT (Poesio and Traum 1997; Poesio and Muskens 1997; Matheson et al. 2000; Poesio and Rieser 2010) is a theory of dialogue semantics and dialogue interpretation developed to explain how utterances are incrementally interpreted in dialogue, considering both their semantic impact and their impact on aspects of dialogue interaction traditionally considered as outside the scope of semantic theory. Much like SDRT (Asher and Lascarides 2003), PTT is a dynamic theory of language interpretation based on DRT (Kamp and Reyle 1993), hence designed to formalize the linguistics of anaphora and reference; but it incorporates ideas about conversation and the construction of the common ground from the work of Clark (1996) and from Situation Semantics (Barwise and Perry 1983; Cooper 1996; Ginzburg To Appear). In this section we briefly introduce the aspects of the theory that are relevant for our discussion of incremental interpretation in dialogue; in the next section we will discuss specifically reference and anaphoric interpretation. For more details on PTT, including a complete fragment for German, see (Poesio and Rieser 2010).

3.1 Compositional DRT

PTT is implemented in Compositional DRT, a reconstruction due to Muskens (1996) of DRT in terms of a standard type logic to which two new types have been added: the type of **discourse referents** π and the type of **states** s. Discourse referents are used to model the dynamics of context in the same way as they are used in DRT, i.e., in the sense that each noun phrase introduces a new discourse referent. States are used to model contexts themselves, and the way they are modified by natural language sentences; they are the object-language equivalent of the assignments used to formalize the semantics of DRSs in (Kamp and Reyle 1993). This dynamics is mediated by DRSs, which in Muskens' type logic are relations between states. A function $v : \pi \rightarrow (s \rightarrow e)$ provides the mapping from discourse referents and states to entities, in the sense that $v(x)(i)$ specifies the 'value' of discourse referent x at state i.

Muskens specifies a translation for all DRT constructs in terms of this type logic. The most important translations are for DRS conditions—general relation and equality predicate **is** —DRSs, and DRS composition, as follows:

(8) a. $R\{x_1 \ldots x_n\}$ is short for $\lambda i.R(v(x_1)(i), \ldots v(x_n)(i))$
b. x_1 **is** x_2 is short for $\lambda i.v(x1)(i) = v(x2)(i)$
c. $[x_1 \ldots x_n|\phi_1 \ldots \phi_m]$ is defined as $\lambda i\lambda j(i[x_1 ... x_n]j \wedge \phi_1(j) \ldots \wedge \phi_m(j))$ where $i[x_1 ... x_n]j$ is short for i and j differ at most over $[x_1 \ldots x_n]$.
d. K;K' is defined as $\lambda i\lambda j(\exists k K(i)(k) \wedge K'(k)(j))$

For example, the type-logic translation of the DRS in (9a) is shown in (9b).

(9) a. $[x, w, y, z, s, s'|\textbf{engine}(x), \textbf{Avon}(w), s : \textbf{at}(x, w), \textbf{boxcar}(y),$
$s' : \textbf{hooked-to}(z, y), z$ **is** $x]$
b. $\lambda i\lambda j i[x, w, y, z, s, s']j \wedge [\textbf{engine}(x)](j) \wedge [\textbf{Avon}(w)](j) \wedge [s : \textbf{at}(x, w)](j) \wedge [\textbf{boxcar}(y)](j) \wedge$
$[s' : \textbf{hooked-to}(z, y)](j) \wedge v(z)(j) = v(x)(j)$

3.2 The Discourse Situation

PTT is an **information state** theory of dialogue (Larsson and Traum 2000; Stone 2004; Ginzburg 2011) in which the participants in a conversation maintain an information state about the conver-

sation consisting of private information together with a conversational score including 'grounded' (Clark 1996) and semi-public information. In PTT, as in Situation Semantics, the conversational score consists of a record of all actions performed during the conversation, i.e., what in Situation Semantics is called the **discourse situation** (Barwise and Perry 1983; Ginzburg To Appear). According to this view, the common ground in an ordinary conversation does not consist only of the content of assertions, but it is a general record of actions the actions that were performed, including actions whose function is to acquire, keep, or release a turn, to signal how the current utterance relates to what has been said before, or to acknowledge what has just been uttered. (Bunt (1995) called these actions **dialogue control** acts.) The discourse situation also contains information about non-verbal actions such as pointing.

Poesio and Traum (1997) argued that the discourse situation-oriented view of the conversational score from Situation Semantics could be formalized using the tools already introduced in DRT (Kamp and Reyle 1993)—specifically, in Muskens's Compositional DRT 1996. Speech acts—**conversational events**, in PTT terms—and non verbal actions are treated just like any other event; conversational events and their propositional contents can serve as the antecedents of anaphoric expressions. For instance, Poesio and Rieser (2010) hypothesize that the two directives in (10) (an edited version of two turns from the Bielefeld ToyPlane Corpus) result in the update to the common ground in (11).[2]

(10)
Inst: So jetzt nimmst du eine orangene Schraube mit einem Schlitz
so now you take a orange screw with a slit

Cnst: Ja
OK

Inst: und steckst sie dadurch, von oben, daß also die drei festgeschraubt werden dann
and you put it through from above so that the three get fixed

(11)
$[\ K1.1, up1.1, ce1.1, K2.1, up2.1, ce2.1|$
$K1.1$ **is** $[e, x, x'|$**screw**(x), **orange**(x), **slit**(x'), **has**(x, x'),
e : **grasp**$(Cnst, x)]$,
$up1.1$: **utter**(*Inst*, *"So jetzt nimmst du ... "*),
$sem(up1.1)$ **is** $K1.1$,
$ce1.1$: **directive**(*Inst&Cnst*, *Cnst*, *K1.1*)
generate$(up1.1, ce1.1)$
$K2.1$ **is** $[z, e', s, w, y|z$ **is** x, e' : **put-through**$(Cnst, z,$ *hole1*$)$,
w **is** $wing1, y$ **is** $fuselage1, s$: **fastened**$(w, y)]$,
$up2.1$: **utter**(*Inst*, *"und steckst sie ... "*),
sem$(up2.1)$ **is** $K2.1$,
$ce2.1$: **directive**(*Inst*, *Cnst*, *K2.1*)
generate$(up2.1, ce2.1)]$

2. We name discourse referents as follows: names with a K prefix for DRSs, with a ce prefix for conversational events, with a u prefix for utterances, with a up prefix for phrasal utterances. As far as content is concerned, we use an e prefix for discourse referents denoting events and the last letters of the alphabet $x, x', y, y', z, z', w, w'$ for other types of discourse referents.

(11) records the occurrence of two conversational events, *ce1.1* and *ce1.2*, both of type **directive** (Matheson et al. 2000) whose propositional contents are separate DRSs specifying the interpretation of the two utterances in (10). The contents of conversational events are associated with propositional discourse referents (discourse referents whose values are DRSs) *K1.1* and *K2.1*, as in (Poesio and Muskens 1997) and in a number of other theories of the common ground, most notably SDRT (Asher and Lascarides 2003). It is further assumed in PTT that dialogue acts are generated (Pollack 1986) by locutionary acts (Austin 1962) which we represent here as events of type **utter**.

Non-verbal actions are also viewed in PTT as conversational events, albeit of a different type. So for instance an act of pointing by agent DG would lead to the following update of both agents' information state:

(12) $[pe1.1|pe1.1 : \mathbf{point}(DG, \alpha)]$

where α is what DG is pointing at. (Determining experimentally what is α is the main question addressed by (Lücking et al. To Appear), as discussed in Section 4.4.)

The extent to which speakers take the common ground into account while referring has been challenged by studies such as (Horton and Keysar 1996). Such studies suggest the need to develop a more nuanced theory of the information state maintained by speakers and how it affects conversational behavior than those developed in response to the original work by Clark and colleagues (Clark and Marshall 1981). And indeed, one of the key differences between PTT and other theories of the common ground that build on DRT is that PTT includes an explicit account of the process by which information becomes part of the common ground, based on the theory of grounding proposed by Traum (1994) and on a theory of the information state in conversations developed in (Poesio and Traum 1997; Matheson et al. 2000; Poesio and Rieser 2010) according to which the information state involves a combination of public, semi-public and private information. In this paper however, for reasons of simplicity, we will omit any discussion of the information state and grounding, and ignore the interaction between incremental interpretation and grounding; see (Poesio and Rieser 2010) for a discussion.

3.3 The Ingredients of Incrementality, I: Micro Conversational Events

It is assumed in PTT that the conversational score is incrementally updated whenever a verbal or non-verbal event is perceived (Poesio 1995a). In particular, each word incrementally updates the discourse situation with a locutionary act of type **utter** and with syntactic expectations about the occurrence of more complex utterances as hypothesized in Lexicalized Tree Adjoining Grammar (LTAG) (Schabes 1988; Abeille and Rambow 2000), that lends itself to a very natural account of the process by which syntactic interpretations are constructed incrementally (Sturt and Crocker 1996).

For instance, an utterance of definite article *the* results in the conversational score being updated with the occurrence of an utterance u_{Det} of syntactic category Det (a **micro conversational event** (MCE) (Poesio 1995a)) and with the expectation that this utterance will be part of an utterance of an NP which will also include an utterance $u_{N'}$ of syntactic category N′.

MCEs are characterized by lexical, syntactic, semantic and discourse information in the form of **features**.[3] One type of syntactic information about MCEs is syntactic constituency; every MCE

3. There is a clear relation between MCEs in PTT and signs in theories such as HPSG, see (Poesio To Appear) for discussion.

u that is not the root of a tree has a mother node u'. We indicate this with the notation used by Muskens (2001) to indicate direct subconstituency in his logic of trees:

$$u \uparrow u'$$

We will assume here the additional features of MCEs in Table 1.

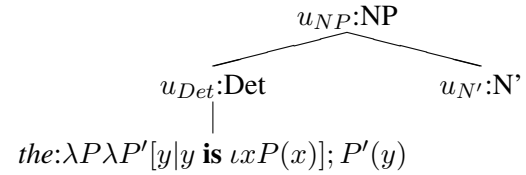
cat	specifying the syntactic category of a MCE
gen	specifying the gender of MCEs
num	specifying the number
sem	specifying its (conventional) semantics
do	specifying the discourse referent introduced by the NP

Table 1: Features of MCES.

The lexical semantics of words that update the discourse model and of anaphoric expressions is as proposed in Compositional DRT (Muskens 1996), according to the grammar fragment discussed in (Poesio and Rieser 2010). The **sem** value of phrasal utterances is obtained compositionally via defeasible inference rules that by default assign, for instance, to an utterance of an NP like u_{NP} above the conventional semantics **sem**(u_{NP}) resulting from the application of **sem**(u_{Det}) to **sem**($u_{N'}$), but that can be overridden e.g., in the case of metonymy or as in the case of anaphoric expressions, as discussed below (Poesio and Traum 1997; Poesio To Appear; Poesio and Rieser 2010).

We will mostly encode the information associated with MCEs in the compact format illustrated by the following example, representing the update resulting from observing an utterance of determiner *the* and by the following lexical access.

(13) [$u_{Det}, u_{NP}, u_{N'}|$



]

3.4 The Ingredients of Incrementality, II: Defeasible Reasoning

The evidence on incremental interpretation and parallel hypothesis generation discussed in section 2 suggests that utterance interpretation is a form of defeasible inference in which competing hypotheses are activated, one of which is rapidly selected, whereas the other ones are discarded (Poesio 1994, 1995a,b, To Appear). This view is also taken in SDRT for aspects of interpretation such as intention recognition or anaphora resolution; in PTT it is assumed that all aspects of utterance interpretation are defeasible, from lexical access to parsing and semantic composition, as already assumed in Hobbs' interpretation as abduction framework (Hobbs et al. 1993) and in the great majority of recent Computational Linguistics work on disambiguation (Hwang and Schubert 1993; Jurafsky 1996; Asher and Lascarides 2003; Bod et al. 2003; Jurafsky and Martin 2009). A good case can be (and has been) made that the defeasible inferences involved in language interpretation are a form of statistical inference (MacDonald et al. 1994; Jurafsky 1996), and most recent theories of interpretation in Computational Linguistics are of this type. However, it is still an open problem

how to combine the logics used in formal semantics with statistical inference,[4] so PTT follows the more traditional approach adopted by virtually all theoretical approaches attempting to combine a theory of performance with a theory of semantic competence based on formal semantics in using a form of logic to model defeasible inference (Perrault 1990; Hobbs et al. 1993; Hwang and Schubert 1993; Poesio 1994, 1995b; Lascarides and Asher 1991; Asher and Lascarides 2003). Specifically, in PTT interpretation is modeled in terms of Prioritized Default Logic (PDL) (Brewka and Eiter 2000; Horty 2007).

For instance, lexical access is modelled in PTT as a Default Theory—a set of defeasible inference rules (specifically, prioritized default rules) specifying the alternative lexical interpretations accessed by encountering an utterance of a given phonetic form. These alternative lexical interpretations are alternative hypotheses about how to update the discourse situation upon hearing that utterance, where each update adds to the discourse situation the information exemplified by (13)—that is, the lexical and LTAG information about that use of the word form. Such hypotheses about updates are produced by (normal) lexical default rule like the rule **LEX-THE** below specifying one of the lexical interpretations of English definite article *the*. **LEX-THE** states that if an utterance of *the* was observed, and it is consistent to hypothesize that this utterance is to be interpreted as the utterance of the determiner of an NP (we will get to the semantics in a moment), then do so.[5]

$$\frac{U : \textbf{utter}(A, \textit{“the”}) \; : \; [\, U_{NP}, U_{N'} \mid \quad \begin{array}{c} U_{NP} : NP \\ U : Det \quad U_{N'} : N' \\ | \\ \textit{the} \end{array} \quad]}{[\, U_{NP}, U_{N'} \mid \quad \begin{array}{c} U_{NP} : NP \\ U : Det \quad U_{N'} : N' \\ | \\ \textit{the} \end{array} \quad]} \textbf{LEX-THE}$$

Homonyms like *stock* or *bank* are associated with multiple such defaults; when these words are encountered, all the **extensions** of the default theory specifying the current information state (i.e., all the inferential closures of the theory obtained using defaults which are consistent) are immediately computed, and if one has a higher priority than the others, that interpretation is chosen and the others remove; else an ambiguity is detected (Poesio 1996, To Appear). We will at times use the standard simplified notation for normal defaults:

LEX-THE

$$U : \textbf{utter}(A, \textit{“the”}) \quad \Rightarrow \quad [\, U_{NP}, U_{N'} \mid \quad \begin{array}{c} U_{NP} : NP \\ U : Det \quad U_{N'} : N' \\ | \\ \textit{the} \end{array} \quad]$$

4. For preliminary work on the matter, see, e.g., (Hwang and Schubert 1993). More recently, Markov Logic Networks have been proposed for this purpose (Richardson and Domingos 2006).
5. Most of the defaults discussed in this paper are open defaults—i.e., default schemata. We use capital letters to indicate the variables in the open default—in this example, U, U_{NP}, and $U_{N'}$ are all variables.

The PTT view of the interpretive processes that follow lexical access such as syntactic interpretation (parsing) is very much inspired by work in grammatical frameworks like Categorial Grammar (Pereira 1990; Carpenter 1998) in that syntactic interpretation is also viewed as an inferential process. Parsing in PTT is a process during which hypotheses about the results of lexical access combine together in phrasal hypotheses through default inference. Such phrasal hypotheses are viewed as hypotheses about utterances of phrases: e.g., the occurrence of contiguous utterances of syntactic category Det and N results in a phrasal hypothesis about the occurrence of an utterance of category NP. These hypotheses are the result of a second set of defeasible inference rules that encode syntactic competence.

PTT is more unusual in proposing that semantic composition, as well, is a form of defeasible reasoning: i.e., that the semantic value **sem** of utterances corresponding to non-terminal nodes like u_{NP} in the example of *the* is specified by default rules which may compete with other defaults (Poesio To Appear). The original motivation for this hypothesis are data about metonymy, and in particular the theory proposed by Nunberg (1995). Nunberg identifies two types of metonymy: **deferred indexical reference** and **predicate transfer**. We are particularly interested in the second of these, illustrated by the utterance in (14), to be imagined uttered by a customer handing his key to an attendant at a parking lot:

(14) I am parked out back.

Nunberg argues that in this example, *am parked out back* is not interpreted as denoting the predicate that holds of objects that are parked out back, but a predicate that applies to human beings whose car is parked out back. I.e., that two different hypotheses about the interpretation of *am parked out back* compete.

(15) $\lambda y.(\forall x[\textbf{car-of}(y) = x] \rightarrow \textbf{parked-out-back}(x))$.

Nunberg argues that predicate transfer is '... a phrasal phenomenon that works in concert with the process of semantic composition' and is subject to the same constraints; e.g., composition has to apply in a certain order.

His description of semantic composition makes it sound very much like a defeasible reasoning process:

> "... one way of dealing with [these cases] would be to permit transfer to take place independently on any simple or complex predicate or term, and then filter the output via constraints charged with maintaining consistency ..." (p. 121)

The conclusion drawn in (Poesio To Appear) is that semantic composition rules, as well, are prioritized defaults. Nunberg's observations can be explained by hypothesizing that at least two defaults apply in this case to derive the **sem** value of the VP node from the meanings of its constituents: a low-priority one, **Binary Semantic Composition** (**BSC**), assigning to a constituent a meaning on the basis of the meaning of its constituents, and a higher-priority one, **PT-BIN-SEM-COMP**, that applies whenever there is a predicate transfer function g mapping Φ into a predicate Υ (e.g., g could be the transfer function mapping predicates like **parked-out-back** into predicates like (15)).

We won't discuss here **PT-BIN-SEM-COMP** (see (Poesio To Appear) for details), only the latest version of **BSC**, proposed in (Poesio and Rieser 2010), which is a direct implementation of type-driven semantic composition. The default specifies that if U_1, U_2 and U_3 are utterances, U_1 and

U_2 are direct constituents of U_3, and the semantic value of U_1 is a function taking as values objects of the type of the semantic value of U_2, then one hypothesis about the semantic value **sem**(U_3) of U_3 is that it results from the application of the semantic value of U_1 to the semantic value of U_2.

$$\frac{U_1 \uparrow U_3, U_2 \uparrow U_3, \textbf{sem}(U_1) \textbf{ is } \phi_{\langle\alpha,\beta\rangle}, \textbf{sem}(U_2) \textbf{ is } \psi_\alpha \ : \ \textbf{sem}(U_3) \textbf{ is } \phi(\psi)}{\textbf{sem}(U_3) \textbf{ is } \phi(\psi)}\textbf{BSC}$$

(Poesio To Appear) also postulates an additional default for percolating the meaning up in the case of nodes with a single constituent, **Unary Semantic Composition** (**USC**).[6]

$$\frac{U_1 \uparrow U_2, \textbf{sem}(U_1) \textbf{ is } \phi_\alpha \ : \ \textbf{sem}(U_2) \textbf{ is } \phi}{\textbf{sem}(U_2) \textbf{ is } \phi}\textbf{USC}$$

We will show in the rest of the paper that incremental interpretation provides further evidence for the hypothesis that semantic composition is defeasible: specifically, we will see that the defaults that produce hypotheses about the interpretation of referring expressions, called **Principles for Anchoring Resource Situations**, are in fact semantic composition defaults.

3.5 The Ingredients of Incrementality, III: Parallelism and Pruning

As discussed in section 2, the view is taking hold that incremental processing should be explained not in terms of serial interpretation as in Frazier's Garden Path model, but in terms of parallel models in which alternative hypotheses are generated in parallel. These hypotheses are sometimes only entertained very briefly before pruning (as in the cases of lexical access first studied by (Swinney 1979)); in others, these hypotheses survive until the end of the sentence (as in the cases of pronoun interpretation studied by (Corbett and Chang 1983)).[7]

This process of generating multiple hypotheses in parallel is naturally modelled in terms of extension generation over the discourse situation. Language interpretation is initiated when the occurrence of a new utterance u is recorded in the information state. At this initial stage, the interpretation of u is **h-underspecified** in the sense of (Poesio To Appear)—i.e., the discourse situation does not specify the value of **sem**(*u*), or its syntactic properties. We can formally characterize the state of the language processor after observing u in terms of the extensions of a default theory generated by prioritized rules like the ones we have discussed. What is still missing to have a complete account of results such as Swinney's is an explanation of the second crucial ingredient of parallel search theories, pruning: i.e., how the language processor decides which extensions to keep and which ones to throw away.

The PTT account of this is quite simple: at the end of each process of extension generation according to the currently active set of PDL rules, only the extensions with highest priority survive; the other ones get pruned. If this hypothesis is correct, at the end of each round of hypothesis generation the processor may find itself in one of two situations. If there is only one remaining extension, the

6. The actual formulation of the defaults is slightly more complex due to the need to ensure that U_1 is the only child of U_2. We assume binary trees only (Poesio 1994)

7. In other cases yet, multiple hypotheses about the interpretation survive even after end-of-sentence processing: this is what happens in cases of deliberate ambiguity, which is fairly common both in political language and in poetry. We called this situation **perceived ambiguity** in (Poesio 1996).

processor **commits** itself to that hypothesis,[8] as in the simplest cases of lexical access in which all hypotheses but the one with highest priority get pruned. This single extension may then represent either a fully specified interpretation, an h-underspecified interpretation, or a **p-underspecified** interpretation (Poesio To Appear; Poesio et al. 2006)—an interpretation which is underspecified in the sense that a more general sense for the word is chosen, as in the cases of lexical polysemy discussed by (Frazier and Rayner 1990). However, it's also possible that more than one extension remains, because more than one conflicting default inference rule with the same priority was activated. At this point different things may happen. The results by (Corbett and Chang 1983) indicate that in some cases of pronoun resolution the conflicting extensions are kept around until the end of the sentence, but then all but one are pruned at that point. [9]

4. Resource Situations, Anaphora, and Reference

In this section we develop the new treatment of definites and anaphoric expressions in PTT proposed in (Poesio and Rieser 2009), to account for the data about incremental reference interpretation. There are two distinctive aspects in this proposal with respect to the standard treatments of anaphora and reference proposed in DRT and SDRT. The first novel aspect is the adoption of the 'functional' interpretation of definite NPs due to Loebner (1987), obtaining a picture of the interpretation of anaphoric expressions with many points in common with the treatment proposed e.g., in (Chierchia 1995). The second is the idea of accessibility via **resource situations** from Situation Semantics, leading to a unified treatment of anaphoric and deictic reference.

4.1 A Reconstruction of Loebner's Theory of Definiteness

According to Loebner, what all definites have in common is that they are *terms*, i.e., functions (in the sense that a Skolem function is a function) that may take a different number of arguments, but all have a value of type e. Thus, a definite *the P* is licensed either because predicate **P** is semantically functional, as in classical examples like *the king of France*, or because **P** is turned into a function by a modifier, as in *the first point to make is that...*, or because **P** is pragmatically coerced into a function by resolving it. Translated into standard logics,[10] the idea is that proper noun *Jack* denotes the (0-argument) function

$$\iota x.(x = j),$$

whereas the definite description *the pope* would denote the 1-argument function

$$\lambda s \iota x.(x = pope(s)(x)),$$

8. This view that committing to an hypothesis is not simply a matter of computing the extensions of a defeasible theory, loosely inspired by Kyburg's work on acceptance (e.g., (Kyburg 1974)) and Pollock's work on a cognitive architecture based on defeasible reasoning (Pollock 1990) clearly requires development, but embedding defeasible jumping to conclusions in a more explicit account of belief maintenance and revision will be required anyway to account for the process by which interpretations are reanalyzed and repaired, a major issue for theories of incremental interpretation. For a probabilistic account of such processes, see (Jurafsky 1996; Schlangen et al. 2009).
9. Finally, in the cases of perceived ambiguity, the conflicting extensions stay around even after the end of the sentence. In other words, a different sort of pruning seems to take place after the first round of hypothesis generation; this second phase of pruning eliminates some interpretations in the Corbett and Chang cases, but not in the case of perceived ambiguity.
10. Loebner only provided an informal discussion of his theory.

taking a situational or temporal argument s.

As just sketched, Loebner's theory would not account for the dynamic properties of definites. The first aspect of our proposal is to combine Loebner's proposals with the treatment of definites in DRT, allowing definites such as *Jack* or *the chair* to update context. This is done by assigning to *Jack* the CDRT semantics

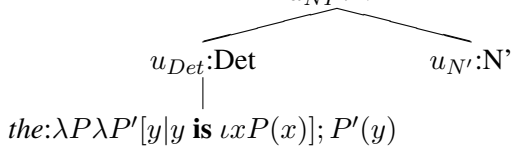
$$\textit{Jack} \Rightarrow \lambda P.([y|y \textbf{ is } \iota x.[|x \textbf{ is } j]]; P(y))$$

i.e., the set of properties of discourse referent y, where y is the unique object that is equal to constant term j denoting Jack (**is** is the equality condition), whereas definite descriptions like *the chair* translate as follows:

(16) $\textit{the chair} \Rightarrow \lambda P.([y|y \textbf{ is } \iota x.[|\textbf{chair}(x)]]; P(y))$

This treatment of definites is implemented in PTT by hypothesizing that a definite article (e.g., English *the*) results in the update to the information state in (17), which combines an LTAG-style prediction of an elementary tree with the CDRT semantics just discussed. The update to the discourse situation in (17) specifies that an utterance u_{Det} of the word *the* has been observed, and that as a result of lexical access, this utterance has been hypothesized to be part of the realization of utterance u_{NP} of an NP part of which has not yet been observed, and has been assigned a **sem** value of $\lambda P \lambda P'([y|y \textbf{ is } \iota x P(x)]; P'(y))$.

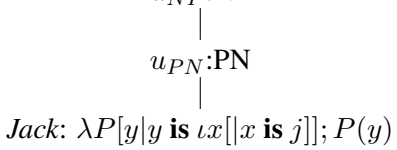
(17) [$u_{Det}, u_{NP}, u_{N'}|$



]

Proper names and pronouns update the discourse situation by adding to it a record of the utterance of a complete NP. The update resulting from proper name *Jack* is as in (18). We'll discuss pronouns in section 4.5.

(18) [$u_{PN}, u_{NP}|$



]

4.2 Resource Situations

In traditional formal semantics, a sharp distinction is made between anaphoric and referential interpretations of expressions such as demonstratives. In an utterance of the sentence *This chair was hand-made by an artisan* accompanied by a pointing gesture to a chair (the demonstration), *this chair* is interpreted as direct reference to the chair. By contrast, in the sentence *Hannes bought a chair in the centre of Rovereto. This chair was hand-made by an artisan*, *this chair* is anaphoric. According to Kaplan (1978), this contrast indicates that demonstrative *this chair* is semantically ambiguous.

The claim that demonstratives are ambiguous has been challenged by Barwise and Perry (1983) and, more recently, by Gundel et al. (1993) in corpus linguistics and by semanticists such as Roberts

(2002). Barwise and Perry proposed that referring expressions like *this chair* in the example above are not ambigous, but depend for their intepretation on a **resource situation**: a situation (in the sense of (Barwise 1989)) containing the object in question that may or may not coincide with the described situation (see also (Ginzburg To Appear)). In the case of *the chair* being used deictically, the resource situation is the visible situation; when it is used anaphorically, it is the described situation. The demonstration is a cue to which resource situation should be used. This proposal was developed in (Gawron and Peters 1990; Poesio 1993; Cooper 1996; Poesio and Muskens 1997). Poesio (1993) proposed a theory of resource situation identification based on prioritized default rules called **principles for anchoring resource situations**, subsequently revised in (Poesio 1994). One of the proposed principles, **PARS1**, produces an hypothesis that (parts of) the visual scene *s* are a possible choice of resource situation when they have been made salient (e.g., as the result of instructions that direct the attention to those parts of the scene).

PARS1 If a speaker uses a referring expression *the P,* the speaker intends the mutual attention of the conversational participants to be focused on the situation *s*, and the visible situation contains an object of type **P**, then the listener may hypothesize that *s* is the resource situation for *the P*.

A second principle, **PARS2**, makes anaphoric reference possible, licensing the choice of the current discourse situation *s* described by **core speech act** (**csa**, (Traum and Hinkelman 1992)) *c* as a possible resource situation for a definite reference *the P* whenever an object of type **P** has been mentioned.[11]

PARS2 If the current discourse topic is the situation or situation kind *s* that includes a discourse marker z of type **P**, a definite NP of the form *the P* may be taken to refer to z.

The proposal in (Poesio 1993) was formulated in terms of Episodic Logic, a logic with situations (Hwang and Schubert 1993). Poesio and Muskens (1997) recast the resource situation proposal in terms of (Compositional) DRT. They proposed that resource situations are contexts—DRSs—and that all anaphoric expressions contain an implicit variable over contexts; it is this variable that supplies the value for the discourse referent.

If we combine these ideas about resource situations with the Loebnerian semantic for definites introduced previously we obtain the single semantic interpretation for definite *the chair* in (19).

(19) $\quad$ *the chair* $\Rightarrow \lambda P'.([y|y \textbf{ is } \iota x.K; [|\textbf{chair}(x)|]]; P'(y))$

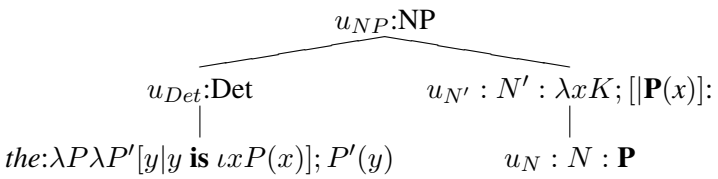
According to the semantics in (19), definite *the chair* gets a presuppositional interpretation requiring the identification of a resource situation K in the context in which an object of type **chair** is particularly salient. Note that K is used presuppositionally—i.e., it is a free variable, just like context variables in Rooth's analysis of focus-sensitive particles (Rooth 1992). The definite gets a deictic interpretation when K is identified with a context specifying (parts of) the visual scene; an anaphoric one when K gets identified with the content of a previous speech act.

Within this new framework, the Principles for Anchoring Resource Situations proposed in earlier PTT work can be reformulated as **coercion rules**: semantic composition rules alternative to the

11. The original formulation of the principle was more general allowing the formulation of such hypotheses also when the previously mentioned object was of type $\mathbf{P}'$ with $\mathbf{P}'$ a lexical prime of **P** (either a synonym or a hyponym) but we will simplify matters here.

default ones discussed earlier, **BSC** and **USC**, that take a non-functional nominal predicate **P** (e.g., *chair*) and turn it into a presuppositional predicate $\lambda x.K; [|\mathbf{P}(x)]$ that is pragmatically functional in the sense of Loebner wrt a resource situation K. For instance, such a coercion rule turns the NP interpretation in (17) into the following one:

(20)



We will consider **PARS1** first, and discuss **PARS2** in the next section. **PARS1** states that the presence of an object *Z* of type **P** in a situation in mutual visual attention K_{mva} is grounds to hypothesize that K_{mva} is the resource situation of a definite description *the P* and Z is the referent of the definite description, i.e., to coerce the interpretation of nominal predicate **P** to the following predicate which clearly is functional in that it only is true of Z:

$$\lambda x K_{mva}; [|\mathbf{P}(x)]; [|x \textbf{ is } Z]$$

This is implemented by formulating **PARS1** as a default rule (of higher priority than **USC** seen before) proposing an alternative specification of the semantic value of the $U_{N'}$ utterance—one in which K_{mva} occurs as resource situation.

In the formulation of **PARS1** below we use a simpler linear notation for representing syntax trees, omitting utterance names where there is no risk of confusion. We also use the notation $K \models \phi$, for K a DRS and ϕ a condition, to indicate that condition K entails ϕ in the sense that all pairs of assignments i, j that verify K must also verify ϕ.

$$\forall i, j K(i)(j) \rightarrow \phi(i)(j)$$

Finally, we adopt a very simple formalization of the notion of mutual visual attention— hypothesizing a distinguished variable $MSOA$ specifying the current mutual situation of attention (see (Grosz 1977); see also (Poesio 1993) for discussion), whose value is constantly updated as mutual attention shifts as the effect of focus shift principles (see section 5). K being the value of $MSOA$ implies mutual belief that K is mutually seen:[12]

$$MSOA \textbf{ is } K_{mva} \rightarrow \textbf{Bel}_{A,B}(A, B, \textbf{\textit{see}}_{A,B}(A,B,K))$$

With this notation **PARS1** is as follows. An utterance of definite *the P* in a discourse situation in which $MSOA$ is K_{mva} and K_{mva} contains an object Z of type **P** leads to hypothesize that Z may be the intended referent of *the P* if it is consistent to assume so.

PARS1
[$U_{NP} : NP$ [Det *the*: $\lambda Q \lambda Q'([y|y$ **is** $\iota x Q(x)]; Q'(y))$] [$U_{N'} : N'$ [N : $\lambda x[|\mathbf{P}(x)]$]]]
$MSOA$ **is** K_{mva}
$K_{mva} \models \mathbf{P}(Z)$
$\Rightarrow$
[$U_{NP} : NP$ [Det *the*: $\lambda Q \lambda Q'([y|y$ **is** $\iota x Q(x)]; Q'(y))$]
[$U_{N'} : N' : \lambda x K_{mva}; [|\mathbf{P}(x)]; [|x$ **is** $Z]$ [N : $\lambda x[|\mathbf{P}(x)]$]]]

12. As mentioned earlier, please refer to (Poesio and Rieser 2010) for a discussion of the information state in PTT and the modalities that qualify its different aspects.

Notice that the uniqueness requirement on *Z*, proposed in (Poesio 1994), has been dropped. This is because in case K_{mva} contains more than one object of type **P** multiple competing hypotheses—i.e., multiple extensions of a default theory with the same priority—would be generated, and as a result, the processor cannot commit to any of them in the sense discussed in section 3.5. This prediction is confirmed by the results of eye-tracking experiments in which multiple objects in the visual situation are briefly considered and maintained until a single interpretation can be obtained (Tanenhaus and Trueswell 2005).

Notice also that the formulation of **PARS1** above only works, stricly speaking, for definite descriptions. This is in keeping with the assumption that distinct interpretation processes apply to each type of referring expression, widely shared among linguists (Gundel et al. 1993), psycholinguists (Garrod 1994) and computational linguists (Sidner 1979; Passonneau 1993; Hoste 2005; Poesio and Kabadjov 2004). We'll make the simplifying assumption in this paper that demonstratives and definite descriptions have the same semantics, and they only differ in that the **PARS3** principle governing interpretation via pointing proposed by Poesio and Rieser (2009) and discussed later in this section only applies to demonstratives, whereas versions of both **PARS1** and **PARS2** apply to both (modulo the triggering condition). The semantics we propose for pronouns however is different, as are the interpretation principles; we'll get back to pronouns after discussing anaphoric accessibility.

4.3 Anaphoric Accessibility via Resource Situations

Before discussing **PARS2** we need to address two apparent problems with the account of incremental reference in discourse situations introduced so far.

The first of these problems is an issue for all theories of anaphoric interpretation that do not make the simplifying assumption that discourse structure is completely flat. Under the anaphoric accessibility rules of DRT, one would conclude that viewing the common ground as a discourse situation leads to the prediction that anaphoric antecedents introduced in previous core speech acts are not accessible, because they are in the scope of the (intentional) operators. Thus for example the antecedent for the pronoun *sie* in (10) (repeated below for convenience), the orange screw with a slit introduced in the first utterance, would be expected to be inaccessible under the view of the common ground in (11), as the screw would be included in proposition *K1.1* (the content of the first speech act *ce1.1*) whereas the pronoun would be part of DRS *K2.1* (the content of the second speech act, *ce2.1*).

(10) Inst: So jetzt nimmst du eine orangene Schraube mit einem Schlitz
so now you take a orange screw with a slit

Cnst: Ja
OK

Inst: und steckst sie dadurch, von oben, daß also die drei festgeschraubt werden dann
and you put it through from above so that the three get fixed

(11) [$K1.1, up1.1, ce1.1, K2.1, up2.1, ce2.1|$
$K1.1$ **is** $[e, x, x'|$**screw**(x), **orange**(x), **slit**(x'), **has**(x, x'),
e : **grasp**$(Cnst, x)]$,
$up1.1$: **utter**(*Inst*, *"So jetzt nimmst du ... "*),
$sem(up1.1)$ **is** $K1.1$,

$ce1.1$: **directive**(*Inst&Cnst*, *Cnst*, *K1.1*)
generate($up1.1, ce1.1$)
$K2.1$ **is** $[z, e', s, w, y | z$ **is** x, e' : **put-through**(*Cnst*, z, *hole1*),
w **is** $wing1, y$ **is** $fuselage1, s$: **fastened**$(w, y)]$,
$up2.1$: **utter**(*Inst*, *"und steckst sie ... "*),
sem($up2.1$) **is** $K2.1$,
$ce2.1$: **directive**(*Inst*, *Cnst*, *K2.1*)
generate($up2.1, ce2.1$)]

But as we said, an explanation for this apparent problem has been available for many years. As argued by Reichman (1985); Grosz and Sidner (1986); Webber (1991); Asher and Lascarides (2003), and others, accessibility in dialogue depends on discourse structure: the antecedents which are accessible are those introduced by utterances belonging to the same discourse segment. In the formulation of Grosz and Sidner, discourse structure depends on intentional structure: utterance U_1 belongs to the same discourse segment as utterance U_2 if the discourse intention of U_2 **satisfaction-precedes** the discourse intention of U_1, whereas it belongs to a subordinate segment if its discourse intention is **dominated** by the discourse intention of U_2. This account was developed most extensively in SDRT (Asher and Lascarides 2003), in which the **VERIDICALITY** axiom ensures that proposition K_1 provides the context for the interpretation of proposition K_2 whenever the speech act with content K_1 is related by one of a small number of discourse relations to the speech act with content K_2.

In (Poesio and Traum 1997), the effect of intentional structure on accessibility was also explained in terms of axioms similar to **VERIDICALITY**, but the formulation of the semantics of definites proposed in this paper, which 'brings the context in' in the form of the resource situation, suggests a solution that does not involve such axioms. We propose instead an explanation for anaphoric accessibility based on a new formulation of the principle governing anaphoric resolution of resource situations, **PARS2**.

This new version of **PARS2** is as follows. Let the referring expression that is to be interpreted, U_{NP}, be a constituent of utterance U, and let U generate a core speech act C', jointly performed by conversational participants A and B. Let C be a core speech act also jointly performed by A and B with content K_{dt} —we indicate this using the notation

$$C : \textbf{csa}(A, B, K_{dt})$$

—and let C **dominate** or **satisfaction-precede** core speech act C'. We use the notation

$$\textbf{accessible}(C, C')$$

to indicate that C either dominates or satisfaction-precedes C' in the sense of (Grosz and Sidner 1986) (see (Poesio 1993, 1994; Poesio and Traum 1997) for details). Then **PARS2** hypothesizes that content K_{dt} is the resource situation for definite U_{NP}.

PARS2
$[U_{NP} : NP$ [Det *the*: $\lambda P \lambda P'([y | y$ **is** $\iota x P(x)]; P'(y))]$ $[U_{N'} : N'$ [N: $\lambda x[|\mathbf{P}(x)]]]]$
$U_{NP} \uparrow U$, **generates**(U, C'),
C : **csa**(A, B, K_{dt}), **accessible**(C, C'), $K_{dt} \models \mathbf{P}(Z)$
$\Rightarrow$
$[U_{NP} : NP$ [Det *the*: $\lambda P \lambda P'([y | y$ **is** $\iota x P(x)]; P'(y))]$
$[U_{N'} : N' : \lambda x K_{dt}; [|\mathbf{P}(x)]; [|x$ **is** $Z]$ [N : $\lambda x[|\mathbf{P}(x)]]]]$

Notice that this proposal amounts to the claim that there are *two* separate attentional structures: one depending on visual attention (implemented here in terms of the $MSOA$ situation) and one depending on accessibility.[13]

There is still an open issue with the current proposal: how can antecedents become accessible in the sense just discussed during incremental interpretation of anaphoric expressions, when the illocutionary force of the utterance to which the anaphoric expression belongs may not yet have been detected? We postpone discussing this issue to the next section. In the rest of this section we complete the discussion of demonstratives and introduce our treatment of pronouns.

4.4 Demonstratives and Pointing

Kaplan (1978) did not actually propose that all referring expressions are ambiguous, only demonstratives; and he did not view all references to objects in the visual situation as directly referring, only those cases which expressed a *demonstration*, usually by pointing. In this section we will see that even the data about demonstratives accompanied by pointing do not require stipulating an ambiguity, reprising the arguments from (Poesio and Rieser 2009).

The main claim of Poesio and Rieser is that the findings from Lücking et al. (To Appear) suggest that pointing is just another way for anchoring resource situations. Using a marker-based optical tracking system, Lücking et al. (To Appear) measured in detail the precision with which the pointing cone projected by an index finger or gaze (Lücking et al. 2010; Pfeiffer 2010) uniquely identifies an object in a visual scene. They concluded that pointing is fuzzy: in most demonstrations the projected ray fails the target. This led them to suggest what they called the **INF** heuristic:

INF (INF) An object is referred to by pointing only if

1. the object is intersected by the pointing cone and
2. the distance of this object from the central axis of the cone is less than any other object's distance within this cone.

INF succeeds in 96% of the cases, which led Poesio and Rieser to formulate what they called **Strong prag hypothesis**:

Strong Prag Hypothesis A pointing gesture refers to the one object selected by an appropriate inference from the set of objects covered by the pointing cone extending from the index finger.

This led Poesio and Rieser to introduce a third principle for anchoring resource situations, **PARS3**—an implementation of the **INF** heuristic. The formulation of the principle proposed here (a slight variant of that proposed in our earlier paper), in addition to coercing the resource situation for the

13. As one of the anonymous reviewers pointed out, this version of anaphoric accessibility may seem overly restrictive in that it doesn't account for cases in which the antecedent of an anaphoric expression is cumulatively constructed out of referents introduced in separate utterances. E.g., in *K1: Mary (x) entered the store. K2: Soon after John (y) reached her. THEY (x+y) were looking for a present for Susan*, the antecedent for *THEY* is the sum of John and Mary, introduced in separate DRSs K1 and K2. We argue that such cases are not particularly problematic but do require to impose constraints like Asher and Lascarides' VERIDICALITY, imposing that interpretation of x in K1 and K2 is consistent. An alternative explanation proposed in (Poesio 1994) was to stipulate the construction of a propositional structure out of the content of the single utterances as proposed by Webber (1991).

demonstrative U_{NP} to the area of the visual situation $K_{pointing}$ that is covered by the pointing cone generated by pointing act P that temporally overlaps with U_{NP}, also proposes as interpretation for the demonstrative NP the object Z which is closest to the pointing axis of the cone. (We won't get here into the best formulation of the notion of 'nearest to' and simply hide the details in a function **nearest-to**; we also assume that every pointing action P has a 'pointing axis' again without entering into any details.)

PARS3
$[U_{NP} : NP$ [Det *this*: $\lambda P \lambda P'([y|y$ **is** $\iota x P(x)]; P'(y))]$ $[U_{N'} : N'$ [N: $\lambda x[|\mathbf{P}(x)]]]]$,
P : **point**$(A, K_{pointing})$, **overlaps**(U_{NP}, P), $K_{pointing} \models \mathbf{P}(Z)$,
Z **is nearest-to**(**pointing-axis**(P)),
$\Rightarrow$
$[U_{NP} : NP$ [Det *this*: $\lambda P \lambda P'([y|y$ **is** $\iota x P(x)]; P'(y))]$
$[U_{N'} : N' : \lambda x K_{pointing}; [|\mathbf{P}(x)]; [|x$ **is** $Z]$ [N : $\lambda x[|\mathbf{P}(x)]]]]$

Notice that as formulated **PARS3** only applies to demonstratives with *this*. We believe a version of the default may exist for demonstratives with *that*, but probably not for definite descriptions.

4.5 Resource Situations and Pronouns

Up until now we have only been concerned with full nominals. We conclude this section by discussing pronouns, beginning with their semantics.

The resource situation idea suggests the following about pronouns. It has often been argued that, syntactically, pronouns in English are like determiners. The translation proposed for pronouns such as German *sie* in (21) makes pronouns behave semantically like determiners, as well.

(21)
u_{NP}:NP
|
u_{Pro}:Pro
|
sie:$\lambda P \lambda P'([y|y$ **is** $\iota x K; P(x)]; P'(y))$

This translation is based on the idea that whereas the definite article may be licensed by a semantically functional, but non anaphoric, predicate, pronouns must always be pragmatically licensed—i.e., there must be some highly salient resource situation K containing a highly salient object. Furthermore, pronouns require a contextual property restricting the interpretation of the referent y: resolving a pronoun amounts to identifying such restriction. One obvious candidate is an identity property—i.e. a property of the form

$$\lambda w([|w \textbf{ is } z])$$

where z is a discourse entity. According to the treatment just sketched, resolving *sie* in (10) involves identifying the content of the first directive in (11), *K1.1*, as the resource situation for the pronoun, and discourse entity z as the antecedent (i.e., applying the result to the identity property $\lambda w([|w$ **is** $z])$.

As said above, our theory of anaphoric interpretation is based on the assumption that the interpretation rules—the Principles for Anchoring Resource Situations—depend very much on the form of the referring expression. We assume therefore that the interpretive steps just discussed are the result of a principle very much like **PARS2**, but which applies specifically to pronouns. We call

the principle **PARS2**$_{pro}$. In first approximation, here is a version of the principle that is exactly as **PARS2** except that is triggered by the occurrence of a pronoun instead of by the occurrence of a definite description. This version of the principle generates a semantic interpretation for any pronominal form PRO which is part of the performance of an utterance U generating core speech act C' such that an antecedent for PRO is available as part of the content of speech act C accessible from C'.

PARS2$_{pro}$
[$U_{NP} : NP$ [Pro PRO: $\lambda P \lambda P'([y|y$ **is** $\iota x K; P(x)]; P'(y))$]]
$U_{NP} \uparrow U$, **generates**(U, C'),
C : **csa**(A, B, K_{dt}), **accessible**(C, C'), $K_{dt} \models$ **P**(Z)
$\Rightarrow$
[$U_{NP} : NP : \lambda P'([y|y$ **is** $\iota x K_{dt}; P(x)]; [y$ **is** $Z]; P'(y))$
[Pro PRO: $\lambda P \lambda P'([y|y$ **is** $\iota x K; P(x)]; P'(y))$]]

This principle produces the interpretation in (21′).

(21') u_{NP}:NP:$\lambda P'[y|y$ **is** $\iota x K 1.1; [|x$ **is** $z]]; P'(y)$
u_{Det}:Det
sie:PRO:$\lambda P \lambda P'([y|y$ **is** $\iota x K; P(x)]; P'(y))$

We will propose a revised formulation of this Principle, including also an agreement check, in section 5.4, after presenting our view of how surface anaphoric expressions like pronouns are interpreted.

5. Accounting for the Evidence about Incremental Reference Resolution

In this section we discuss how the proposals about modelling incremental language processing as a process of defeasible inference and about the semantics of anaphoric expressions discussed in the previous section can account for the evidence about the incremental interpretation of anaphoric expressions coming from the psycholinguistics literature.

5.1 Basics: Incremental Resolution of References to the Visual Scene

We begin by showing how the proposed theory accounts for the fundamental results concerning incrementality in reference to objects in the visual world from Tanenhaus et al. (1995) and Eberhard et al. (1995). Let us consider the two types of visual world situation studied by Tanenhaus and colleagues and shown in Figure 1. In the situation on the left there is only one apple; in the situation on the right there are two. Subjects looking at either visual situation hear the instruction "Put the apple on the towel in the box."

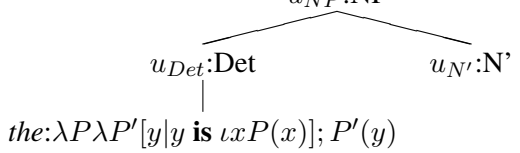
The first word utterance in the instruction, of verb *put*, leads to updating the discourse situation first by recording the occurrence of an utterance of word "put", and then by the results of lexical access—which, in the LTAG framework adopted here, means predicting the utterance not only of a VP, but of a whole sentence, as discussed in detail in (Poesio and Rieser 2010). To keep the syntactic trees in this section manageable we will therefore omit representing this part of the phrasal structure. Let us instead consider in more detail what happens according to PTT when processing the next words in the instruction, forming the referring expression *the apple*.

Perceiving an utterance of determiner *the* results in the discourse situation being updated with the observation of an utterance of determiner *the*, as in (22).

(22) [u_{the} : **utter**$(A, \text{“}the''\text{)}$]

This update leads in turn to the parallel activation of all lexical access defaults associated with word form *the*, and in particular lexical default **LEX-THE** discussed in section 3.4. This in turn leads to the update in (17) here repeated for convenience. (We are not concerned here with wordsense disambiguation, but anyway we can assume it's not a major issue in the case of this utterance.)

(17) [$u_{Det}, u_{NP}, u_{N'}|$



]

Next (or while the interpretation inferences activated by *the* are taking place), perceiving the utterance of noun *apple* leads to the update in (23), which in turn leads again to lexical access, i.e, to the parallel activation of all lexical defaults, and to the selection of the interpretation in (24) through wordsense disambiguation processes which are not our concern here.[14]

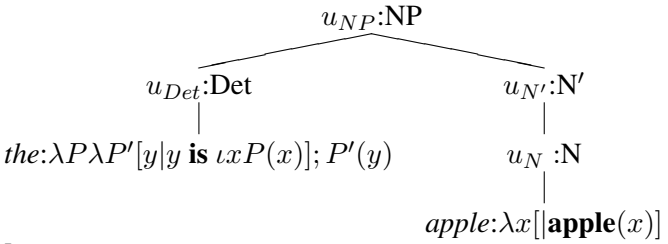
(23) [u_{apple} : **utter**$(A, \text{“}apple''\text{)}$]

(24) [$u_N|$

]

Parsing then results in the interpretation in (25), through one of the basic LTAG operations, substitution of u_N into u_{NP}.

(25) [$u_{Det}, u_{NP}, u_{N'}|$



]

In both visual world scenarios in Figure 1 the entire visual world—called here K_{visual}—is in the mutual focus of attention

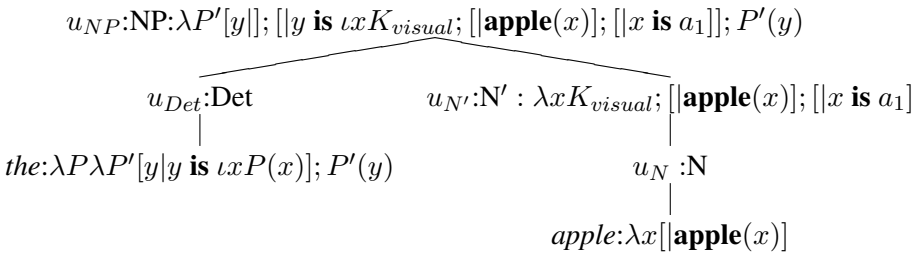
$$MSOA = K_{visual}$$

so that the subject can use **PARS1** to assign K_{visual} as the resource situation for the definite. In the case of a single apple—let us call that a_1 —**PARS1** can only be applied once, producing the single hypothesis in (26). This interpretation is fully specified: the discourse situation provides syntactic

14. Two senses are listed for *apple* in WordNet—the fruit sense and the tree sense—but this is presumably a case of polysemy which would be handled in PTT by assuming a p-underspecified lexical interpretation, see (Poesio To Appear).

and semantic interpretations for all phrasal and lexical MCES. The processor can therefore commit to the interpretation in (26), which results in the concentration of fixations on the target word observed in such situations (see e.g., the introduction to such results at pages 13–14 of (Tanenhaus and Trueswell 2005)). (Committing to this hypothesis also prevents further attachments, as discussed below.)

(26)



In (26), the interpretation for u_{NP} is the set of properties that hold of discourse referent y such that y is the only object in K_{visual} that is an apple and is equal to a_1. By contrast, in the situation illustrated on the right of Figure 1, **PARS1** can be applied twice to produce two hypotheses concerning the **sem** value of utterance $u_{N'}$ —as in (26), and the interpretation which is identical to the one (26) in all respects except that the cohort apple is chosen (let us call this second apple a_2). This leads to the fixations being divided between the two apples.[15] However both of these interpretations satisfy the uniqueness requirement on the interpretation of the definite. As a result, subjects cannot commit to a single extension, and therefore cannot assign an interpretation to the NP utterance u_{NP}, as discussed in section 3.5, and therefore have to backtrack, so that when the rest of the instruction, ... *on the towel in the box*, comes in, hypothesis (25) is still open to modification, which results in the lack of garden path effect in this case, as discussed in section 5.3.

5.2 Incremental Establishment of Referential Domains

The key difference between the resource situation view of domain restriction and standard theories of quantifier domain restrictions such as those proposed, e.g., in (Partee 1991; Rooth 1992), in which any contextually salient property **P** can serve to restrict the domain of a quantifier, is the idea that the domain is restricted to the objects of a situation—a spatially and possibly temporally limited set of objects. In our original work (Poesio 1993, 1994) this stronger formulation of domain restriction (at least for definite descriptions), inspired by Grosz (1977), was motivated by the fact that restriction domain shifts in the TRAINS dialogues appeared to be tied in with locations on the map that participants in the experiments are looking at, which represents a highly simplified 'world' with a few towns represented as circles and connected by lines representing railways—i.e., moving a train to a town seemed to restrict the domain of interpretation to the area of the map around that town. This led to the formulation of the hypothesis that the TRAINS world had a structure, in the sense that at the very least each landmark in the world identified a sub-situation that could serve as $MSOA$; it was also possible that larger sub-situations could be identified. And whereas Grosz

15. Numbers of fixations in visual world studies are computed across a number of subjects, which, as pointed out by one of the reviewers of this papers, leads to the question whether each subject divides her/his attention between the extensions, or whether half of the subjects fixate on one interpretation whereas the other half fixate on the second. The matter clearly requires more empirical evidence, but judging from our experience with ambiguity perception (e.g., (Poesio and Artstein 2005)), we feel that the second explanation is more likely, and that individual subjects choose to fixate between the two equally likely candidates randomly as discussed in (Poesio 1994).

(1977) had identified focus shifting principles tied to the structure of the task, we identified a new, spatially related (visual) focus shifting principle that we called **FOLLOW-THE-MOVEMENT**:

FOLLOW-THE-MOVEMENT Part of the intended effect of an utterance instructing an agent to move an object from one location to another is to make the terminal location of the movement the new mutual situation of attention.

One of the great opportunities offered by the development of the visual-world methodology was the possibility to investigate in a proper empirical fashion the relation between shifts in the visual focus and reference interpretation, and indeed a key line of research in this type of work has been the study of incremental focus shifting, under the name **rapid restriction of referential domains** (Chambers et al. 2002; Brown-Schmidt et al. 2005).

The experiments by Brown-Schmidt *et al.* discussed in section 2 are the experimental setting closest to that of the TRAINS dialogues. Although the experimenters did not directly test specific focus-shift principles, the results clearly confirm the hypothesis that spatial landmarks identify sub-situations from the attentional point of view. The preliminary results of a more direct test of **FOLLOW-THE-MOVEMENT** just undertaken in our lab (in preparation), and (more indirectly) of the generation experiments in (Zender 2010), also appear to confirm the existence of that focus shift principle. We hypothesize therefore that at least in the simplified type of visual scenes used in visual world experiments or in the TRAINS dialogues, each landmark l identifies a visual sub-situation K_l. Having made this assumption, **FOLLOW-THE-MOVEMENT** translates into the following default: if A intends A, B to move object C to landmark L, A also intends sub-situation K_L to be the new $MSOA$.

FOLLOW-THE-MOVEMENT
$\textbf{Int}_A(\textit{move}(\{A,B\},C,L)), \Rightarrow \textbf{Int}_A(MSOA\ \textit{is}\ K_L)$

The data on referential domain restriction by Chambers *et al.*, however, seem to indicate that the interpretation domain can also be restricted not according to spatial location, but according to what Brown-Schmidt *et al.* call **task compatibility**: after hearing *Pick up the cube. Put it in ...*, attention is focused on the set of containers into which the cube can fit. This suggests that a more general formulation of domain restriction principles is needed than we proposed in our previous work, one in which the resource situation need not be spatially defined. We claim that the new formulation of resource situations in terms of DRSs proposed in this paper is exactly what is needed and covers these types of referential domain restriction as well. Specifically, we propose that upon hearing the utterance *Pick up the cube. Put it in ...*, the discourse situation is updated by introducing a new resource situation $K_{fit-cube}$ thus defined:

(27) $[\ K_{fit-cube} |\ K_{fit-cube}\ \textbf{is}\ [x, y, s | x\ \textbf{is}\ \iota z.\textbf{pick-up}(subj, z), \textbf{container}(y), s : \textbf{fit-in}(x, y)],]$

and this new resource situation then becomes the $MSOA$, ensuring that $\textbf{PARS2}_{pro}$ would only suggest those objects as antecedents of following references to containers.

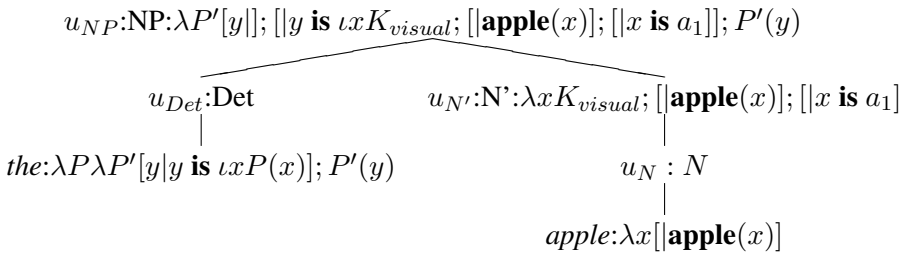
5.3 The Interaction between Reference and Parsing

Next, let us discuss how the proposal presented in this paper can account for the interaction between reference and parsing studied in (Crain and Steedman 1985; Altmann and Steedman 1988;

Tanenhaus et al. 1995; Spivey et al. 2002). Let us consider again the two types of experimental settings constrasted in the study by Tanenhaus and colleagues (Figure 1, left and right) which we just discussed focusing exclusively on reference interpretation.

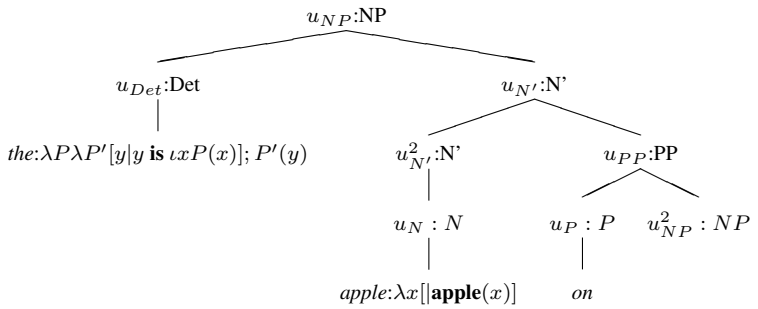
Let us begin again with the visual world context on the left, in which there is only one apple. As just discussed, in this situation the processor can use **PARS1** to choose the current $MSOA$, K_{visual}, as the resource situation, and choose the only apple in K_{visual}, that we called a_1, to produce the single hypothesis in (26), repeated below for convenience.

(26)



As discussed above, this interpretation is fully specified: the syntactic and semantic interpretations of each phrasal and lexical MCE are fixed by the discourse situation. The subject can therefore commit to the interpretation in (26), preventing further attachments. So when the subject hears next the utterance of a PP, *on the towel*, the only available interpretation is as an argument of *put*, leading to a garden path.

By contrast, in the case of the visual situation on the right of Figure 1, the hypotheses produced using **PARS1** ((26) and the interpretation which is identical to (26) in all respects except that apple a_2 is chosen) could not be committed to as they did not satisfy the uniqueness restriction imposed by the determiner, and therefore subjects have to backtrack to hypothesis (25). This means that when the next part of the instruction, ... *on the towel*, comes in, this hypothesis is still open to modification—in fact, it requires the meaning of $u_{N'}$ to be restricted in order to find a discourse referent satisfying the uniqueness condition. We argue that this requirement is what makes adjunction of ... *on* ... to $u_{N'}$ preferred over substitution as second argument of *put*. As a result of this adjunction we obtain:

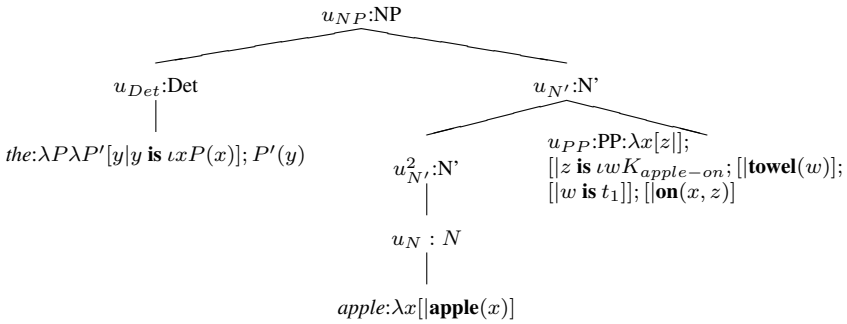


At this point a second definite NP is uttered, *the towel*. A crucial point in need for an explanation about this example is the fact that this definite NP is felicitous in a context in which there are two towels. We claim that this is another case of task compatibility leading to rapid referential domain adaptation, just as in the cases studied by Chambers *et al.* and whose analysis we presented in the previous section. I.e., we claim that upon hearing *on*, the discourse situation is updated by changing the $MSOA$ to a new resource situation containing the objects on which an apple is placed, as follows.

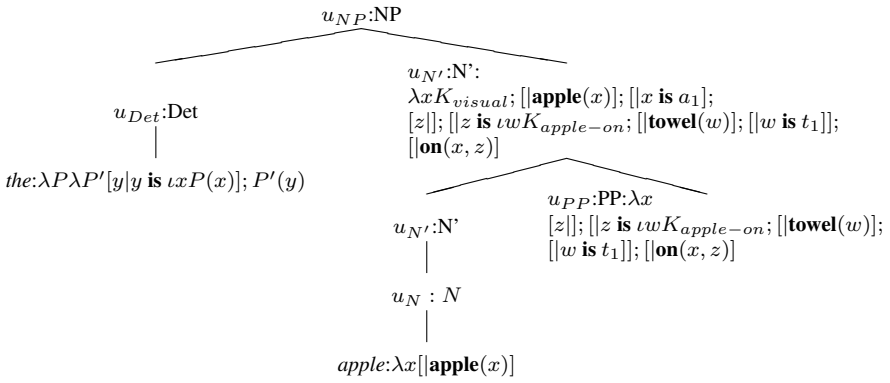
(28) $[K_{apple-on}, MSOA|$
$K_{apple-on}$ **is** $[x, y, s|x$ **is** $\iota z.[|\textbf{apple}(z), \textbf{put}(subj, z, y)], \textbf{object}(y), s : \textbf{on}(x, y)]$,
$MSOA$ **is** $K_{apple-on}$]

Notice that the new resource situation only contains one towel, the towel in the top left quadrant, that we will call t_1. $K_{apple-on}$ can now be chosen as resource situation for *the towel* via **PARS1**; this makes the definite felicitous.[16]

The interpretation resulting from this first application of **PARS1** is shown below. According to this interpretation, *the apple on the towel* gets interpreted as *the apple on the unique towel in the visual scene that has an apple on it*.



As this update makes the meaning of $u_{N'}$ functional, **PARS1** can now apply to choose the visual situation K_{visual} as resource situation for the first definite, identifying a_1 on towel t_1, as shown below.



As this interpretation is fully specified, the subject can commit.

5.4 Incremental Interpretation of Anaphoric Reference via Pronouns

Whereas the experiments by Tanenhaus *et al.* and by Eberhard *et al.* were only concerned with references via full nominals to entities in the visual situation, the visual world methodology has also been shown in experiments such as those by Arnold et al. (2000) to confirm earlier evidence (by, e.g., (Corbett and Chang 1983)) that pronouns are interpreted incrementally, as well.

16. The idea that *the towel* in this case is felicitous in virtue of being interpreted as, essentially, 'the towel that an apple is on' is reminiscent of the interpretation for definites proposed by Webber in her thesis (Webber 1979).

We now discuss how the new version of **PARS2** for pronouns proposed in section 4.5, **PARS2**$_{pro}$, can explain how an interpretation is assigned to the pronouns in the experiments discussed by Arnold *et al.*, repeated here.

(29) a. Donald is bringing some mail to [Mickey / Minnie]
while a violent storm is beginning.

b. He's / she's carrying an umbrella,
and it looks like they're both going to need it.

As we are not concerned with speech act interpretation and discourse structure recognition in this paper, we will just make some assumptions here about the results of these interpretation processes. We believe that a fuller account could be developed building on the detailed analysis of these interpretation processes proposed by Asher and Lascarides (2003) in the SDRT framework, which shares many assumptions with PTT.

The first utterance, (29a), generates a core speech act of type **assert**, that we will call *ce1*. Overall, the update resulting from the first utterance is then as in (30a), where A is the experimenter and B the experimental subject. In this example we have included in the description of the discourse situation, in addition to full utterance $up1$, two of its subconstituents: the micro conversational events $up1.1$ of uttering name "Donald" and $up1.2$ of uttering name "Mickey," in both cases omitting syntactic and semantic information for these MCEs except for their gender. The update resulting from the variant with Minnie instead of Mickey, in (29b), is the same as (29a) except that in this case $x3$ refers to Minnie and the gender of $up1.2$ is feminine.

(30) a. [$K1, up1, ce1, up1.1, up1.2|$
$K1$ **is** $[e1, x1, x2, x3|x1$ **is** $Donald$, **mail**$(x2), x3$ **is** $Mickey$,
$e1$: **bring**$(x1, x2, x3)]$,
$up1.1$: **utter**$(A,$ *"Donald"*$)$,
gen$(up1.1)$ **is** $masc$,
$up1.1 \uparrow up1$,
$up1.2$: **utter**$(A,$ *"Mickey"*$)$,
gen$(up1.2)$ **is** $masc$,
$up1.2 \uparrow up1$,
$up1$: **utter**$(A,$ *"Donald is bringing some mail to Mickey"*$)$,
sem$(up1)$ **is** $K1$,
$ce1$: **assert**$(A, B,$ *K1*$)$
generate$(up1, ce1)$]

b. [$K1, up1, ce1, up1.1, up1.2|$
$K1$ **is** $[e1, x1, x2, x3|x1$ **is** $Donald$, **mail**$(x2), x3$ **is** $Minnie$
$e1$: **bring**$(x1, x2, x3)]$,
$up1.1$: **utter**$(A,$ *"Donald"*$)$,
gen$(up1.1)$ **is** $masc$,
$up1.1 \uparrow up1$,
$up1.2$: **utter**$(A,$ *"Minnie"*$)$,
gen$(up1.2)$ **is** fem,
$up1.2 \uparrow up1$,
$up1$: **utter**$(A,$ *"Donald is bringing some mail to Minnie"*$)$,
sem$(up1)$ **is** $K1$,

$ce1$: **assert**$(A, B, K1)$
generate$(up1, ce1)$]

The pronoun (*He* or *She*) uttered at the beginning of the second utterance (utterance *up2.1*) is interpreted as the beginning of an utterance *up2* generating a second core speech act *ce2* whose precise type we do not know yet.[17] We show the update resulting from *He* in (31a), that resulting from *She* in (31b).

(31) a. [$K2, up2, ce2|$ **utterance**$(up2)$,$ce2$: **csa**$(A, B, K2)$,
generate$(up2, ce2)$,$up2.1$: **utter**$(A,$ *"He"*$)$, $up2.1 \uparrow up2$,
gen$(up2.1)$ **is** $masc$]

b. [$K2, up2, ce2|$ **utterance**$(up2)$,$ce2$: **csa**$(A, B, K2)$,
generate$(up2, ce2)up2.1$: **utter**$(A,$ *"She"*$)$, $up2.1 \uparrow up2$,
gen$(up2.1)$ **is** fem]

According to the theory of resource interpretation anchoring developed in section 4, in the case of references to the visual situation it doesn't matter that the core speech act of whose realization the referring expression is part, or its connection with the rest of the discourse structure, hasn't yet been identifed at the time the referring expression is uttered, because the interpretation of the referring expression only depends on the visual attentional state, as opposed to the linguistic attentional state. On the other hand this does matter in the case of anaphoric references, like pronoun *sie* in (10) or the pronouns in the example under discussion. This is because principles **PARS2** and **PARS2**$_{pro}$, in order to choose the content of a previous core speech act *C* as resource situation, require that core speech act to **dominate** or **satisfaction-precede** the core speech act containing the anaphor. The question is, how can anaphora resolution proceed prior to recognizing the intention behind the utterance being produced?

As in SDRT, and consistent with the general view of interpretation discussed in section 3, discourse structure recognition is viewed in PTT as a defeasible inference process. This means that hypotheses about discourse structure and accessibility are produced before complete information is available. In fact, in PTT it is assumed that such hypotheses tend to be produced very early on the basis of fairly superficial information, and possibly revised later. In cases like the example under discussion, we assume that the accessibility of (the content of) *ce1* is hypothesized before knowing the content of *ce2*—possibly even before knowing its illocutionary force.

There are two possible time points at which this accessibility hypothesis may be produced. It could be produced immediately, on coherence grounds: i.e., in circumstances in which it would seem that a story is being told, simply assume by default that the next utterance is going to tell the next episode in the story, by way of a default that would be like a highly underspecified version of Asher and Lascarides's **NARRATION** (Asher and Lascarides 2003). Alternatively, the accessibility hypothesis might be produced as a byproduct of establishing a preliminary link between the pronoun and one of the antecedents. We will only pursue here this second possibility, as in this way we can also spell out more fully the view of anaphoric processing adopted in PTT. (Anyway, more empirical evidence about the precise time point of discourse structure identification is needed before being able to decide which of the possibilities is more likely, or whether the interpretation results from a combination of the two factors.)

17. As pointed out by one of the reviewers, a full account of incremental processing would also require an account of the process by which the beginning of an utterance is recognized.

The treatment of 'surface' anaphora resolution (Hankamer and Sag 1976) we propose here is based, on the one hand, on the proposal by Garrod (1994) and Garrod and Sanford (1994) that the resolution of these types of anaphoric reference consists of distinct **bonding** and **resolution** stages; on the other, on Centering theory (Grosz et al. 1995; Poesio et al. 2004). According to Garrod and Sanford, in the initial bonding stage a link is made between the anaphoric expression and one or more candidate antecedents in the discourse context, on the basis of superficial information. In the subsequent resolution phase, the link made in the bonding stage is evaluated, recomputed if necessary, and integrated into the semantic interpretation.

What is meant by 'superficial information' has never been spelled out in detail by Garrod and Sanford, but we propose here that the 'superficial level' is the micro conversational events level hypothesized in PTT: i.e., that at least some of the defeasible rules for anaphora resolution establish bonding links between the micro conversational events introducing discourse antecedents. We further propose that these MCEs are the **forward looking centers** (CFs) of Centering, another notion whose linguistic characterization has never been spelled out fully (see also (Poesio 1994, To Appear)).

These bonding links between CFs, in turn, lead to hypothesizing dominance or satisfaction-precedes links between the core speech acts generated by the utterance of which the anaphoric CF is a constituent and the utterance which includes the antecedent CF. This link, finally, enables **PARS2**$_{pro}$—our proposal concerning the 'resolution' stage.

This theory is implemented by assuming, first of all, that the local focus is recomputed after what we will call here **c-utterance**, for 'Centering Utterance', as proposed in Centering theory.[18] This translates into hypothesizing that the discourse situation has a distinguished discourse marker (in the sense of CDRT) called CCU (for 'Current C-Utterance'), whose value changes after every sentence; we also hypothesize that end-of-sentence processes include updating the discourse situation with several statements of the form

$$\textbf{cf-utt}(u, u')$$

indicating that u' is a CF in c-utterance u.

Second, we stipulate that at least some of the mechanisms for interpreting pronouns operate at the surface level: specifically, that (one of) the default rules for pronoun resolution, that we call here **PRO-MATCH**, creates a bonding link between the utterance of a pronoun and one of the CFs of the current c-utterance, provided that their agreement features match. We write **bond**(u,u') to indicate that u' is bonded to u: in the case of this default rule, that the utterance U_{pro} of a pronoun realized in CCU U_{n+1} is bonded to the utterance U_{np} of a CF realized in U_n and which agrees with the pronoun in gender person and number.[19]

$$\frac{\begin{array}{l}\textbf{cat}(U_{pro})\ \textbf{is}\ \text{PRO},\\ \textbf{cf-utt}(U_{n+1},U_{pro}),\\ \textbf{cf-utt}(U_n,U_{np}),\\ CCU\ \textbf{is}\ U_{n+1},\\ \textbf{agr-match}(U_{np},U_{pro})\end{array} \quad : \textbf{bond}(U_{np},U_{pro})}{\textbf{bond}(U_{np},U_{pro})}\ \textbf{PRO-MATCH}$$

18. The notion of 'utterance' is used in Centering to indicate the amount of language after which the local focus gets updated. We identify here 'utterances' in the Centering sense with sentences, on the basis of the results in (Poesio et al. 2004).

19. Although we only propose here this treatment for surface anaphors, we suspect a similar **DD-MATCH** rule may generate anaphoric interpretation hypotheses for definite descriptions on the basis of head similarity.

The establishment of bonding links is one trigger for further inference processes that hypothesize dominance / satisfaction precedes relations between the core speech acts generated by the two utterances, if they haven't already been established by coherence assumptions or by previous intention recognition processes. The following default, **ACC-FROM-BOND**, hypothesizes that conversational action CE_1 is **accessible** in the sense discussed in section 4.3 from conversational action CE_2 if a pronoun that is part of the realization of CE_1 is bonded to a CF that is part of the realization of CE_2.

$$\frac{\begin{array}{l}\textbf{cat}(U_{pro})\ \textbf{is}\ \text{PRO},\\ \textbf{cf-utt}(U_{n+1},U_{pro}),\\ \textbf{cf-utt}(U_n,U_{np}),\\ \textbf{bond}(U_{np},U_{pro}),\\ \textbf{generates}(U_{n+1},CE_2),\\ \textbf{generates}(U_n,CE_1)\end{array} \quad : \ \textbf{accessible}(CE_1,CE_2)}{\textbf{accessible}(CE_1,CE_2)}\ \textbf{ACC-FROM-BOND}$$

After establishing accessibility through **ACC-FROM-BOND**, or possibly through other shallow coherence-inference methods, the resolution stage can begin and **PARS2**$_{pro}$ can be applied to identify the resource situation.

We can finally produce the revised version of **PARS2**$_{pro}$ promised earlier in the paper. Whereas **PARS2** for definites only depends on the existence of an object of the appropriate type in the resource situation, according to **PARS2**$_{pro}$ the identification of an antecedent for a pronoun utterance U_{pro} depends on having established (via agreement matching) a bonding link with a forward-looking center U^1_{NP} in the context. The pronoun is not just resolved to any referent in the content of a context accessible from the core speech act being produced, but to the described object **do** of U^1_{NP}.

PARS2$_{pro}$
[$U_{pro} : NP$ [Pro PRO: $\lambda P \lambda P'([y|y$ **is** $\iota x K; P(x)]; P'(y))]$]
$U_{pro} \uparrow U$, **generates**(U, C),
bond(U_{NP}, U_{pro}),
$U_{NP} \uparrow U'$, **generates**(U', C'),
C : **csa**(A, B, K_{dt}), **accessible**(C', C),
do(U_{NP}) **is** Z, $K_{dt} \models \mathbf{P}(Z)$
$\Rightarrow$
[$U_{pro} : NP\ \lambda P'([y|y$ **is** $\iota x K_{dt}; [|y$ **is** $Z]]; P'(y))$ [Pro PRO: $\lambda P \lambda P'([y|y$ **is** $\iota x K; P(x)]; P'(y))]$]

Let us now return to the data from Arnold and consider first (31a) uttered in the discourse situation following (30b), in which the two CFs are of different gender. As only MCE $up1.1$ (the utterance of "Donald") matches $up2.1$ in gender, **PRO-MATCH** can only produce one hypothesis about the interpretation of $up2.1$: that it bonds to $up1.1$—i.e., to the update in (32). (The same happens when "she" is uttered, with both contexts in (30).)

(32) [|**bond**(*up1.1*, *up2.1*)]

This hypothesis is immediately committed to, resulting in **ACC-FROM-BOND** being triggered. This in turn leads to the following update:

(33) [|**accessible**(*ce1*, *ce2*)]

which in turn triggers **PARS2**$_{pro}$, resulting in the interpretation in (34)

(34) $$\begin{array}{c} u_{NP}\text{:NP}\lambda P'[y|y \textbf{ is } \iota x K1; [|x \textbf{ is } z]]; P'(y) \\ | \\ u_{Det}\text{:Det} \\ | \\ \textit{sie}:\lambda P \lambda P'([y|y \textbf{ is } \iota x K; P(x)]; P'(y)) \end{array}$$

It is not clear from the results of Arnold *et al.* whether the concentration of the fixations on the target starting from around 400msec after the onset of the pronoun is the result of bonding or of resolution; more experimental evidence is needed to resolve the issue.

Let us now consider the case of (30a), in which both $up1.1$ and $up1.2$ match $up2.1$ in gender. As a result, **PRO-MATCH** can be activated in two different ways, producing the two distinct hypotheses in (35)

(35) a. [|**bond**(*up1.1*, *up2.1*)]

b. [|**bond**(*up1.2*, *up2.1*)]

Each of these hypotheses in turn activates **ACC-FROM-BOND**—the updates resulting from this default are however identical (and identical with (33)). Arnold *et al.*'s results are that in case the same-gender target is the first mentioned entity, fixations quickly concentrate on the target, whereas if the target is the second-mentioned entity, the subjects look at both the target and the competitor in the same amount. This situation is reminiscent of the situation with lexical interpretation and scope access discussed in (Poesio 1994, 1996), and suggests that a stronger default than **PRO-MATCH** is at play in the case of first-mention entities. When this default, shown below and that we call **PRO-MATCH-FM**, is triggered, it overrides the weaker **PRO-MATCH**; otherwise a conflict between weaker defaults is obtained, which typically results in a toss-up between the alternatives.

$$\frac{\begin{array}{l}\textbf{cat}(U_{pro}) \textbf{ is } \text{PRO}, \\ \textbf{cf-utt}(U_{n+1}, U_{pro}), \\ \textbf{cf-utt}(U_n, U_{np}), \\ \textbf{first-mention}(U_n, U_{np}), \\ CU \textbf{ is } U_{n+1}, \\ \textbf{agr-match}(U_{np}, U_{pro})\end{array} \quad : \ \textbf{bond}(U_{np}, U_{pro})}{\textbf{bond}(U_{np}, U_{pro})}\ \textbf{PRO-MATCH-FM}$$

(A more elegant formalization would of course be available in a framework with evidence accumulation.)

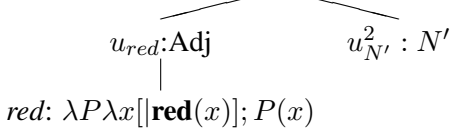
5.5 Reference Interpretation Prior to Hearing a Complete Head Noun

Eberhard et al. (1995) and Allopenna et al. (1998) showed that interpretation processes begin much earlier than discussed until now: they begin as soon as an unambiguous phonetic prefix has been uttered. Allopenna et al. (1998) also showed that in the case of the interpretation of referring expressions, this unambiguous phonetic prefix need not be part of the head noun—in a situation in which there is a single red object, and *click on the red triangle* is uttered, fixations concentrate on that object as soon as adjective *red* has been perceived, without waiting to hear *triangle*. A proper account of the incremental effect of sub-word prefixes would require an implementation in PTT of a theory of sub-word based lexical access such as the cohort model (Marslen-Wilson 1987) or the TRACE model (McClelland and Elman 1986), so we will not attempt that here. We will however discuss how the present proposal accounts for how hearing an adjective affects reference resolution.

The interpretation process resulting from an utterance of *the* is as discussed above, and results again in update (17). Upon hearing *red*, the discourse situation is updated with the expectation of encountering an N', as in (36). After parsing, this interpretation is adjoined to the syntactic interpretation of *the*, resulting in the updated interpretation in (38).

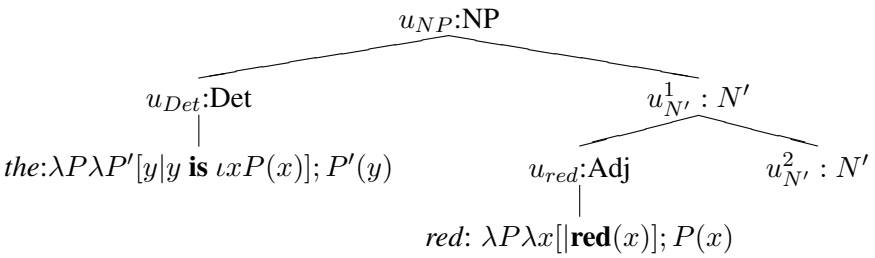
(36) [u_{red} : **utter**$(A, "red")$]

(37) [$u^1_{N'}, u^2_{N'}|$



]

(38) [|



]

The evidence from Eberhard *et al.* suggests that, at least in interpretive contexts like the visual world scenarios, updates like (38) are sufficient to trigger reference resolution. This translates into an hypothesis that a version of **PARS1** exists activated by the observation of an utterance of an adjective with semantic interpretation $\lambda P \lambda x([|\mathbf{P}(x)]; P(x))$. This version of **PARS1**, that we call $\mathbf{PARS1}_{Adj}$, is shown below. This version again requires a visual situation K_{mva} to be in the mutual focus of attention, but unlike the versions of the default proposed earlier, it is sufficient for predicate **P** to hold of object Z, even if the predicate is not the head of an NP. The default updates the discourse situation by restricting the interpretation of the NP through adjoining.

$\mathbf{PARS1}_{Adj}$
[$U_{NP} : NP$ [Det *the*: $\lambda P \lambda P'([y|y$ **is** $\iota x P(x)]; P'(y))$]
[$U^1_{N'} : N'$ [U^1_{Adj} : *Adj*:$\lambda P \lambda x[|\mathbf{P}(x)]; P(x)$]
$U^2_{N'} : N'$]]
$MSOA$ **is** K_{mva}
$K_{mva} \models \mathbf{P}(Z)$
$\Rightarrow$
[$U_{NP} : NP$ [Det *the*: $\lambda P \lambda P'([y|y$ **is** $\iota x P(x)]; P'(y))$]
[$U^1_{N'} : N'$ [U^2_{Adj} :Adj:$\lambda P \lambda x K_{mva}$; [$|\mathbf{P}(x)$]; [$|x$ **is** Z]; $P(x)$
[U_{Adj} :Adj: $\lambda P \lambda x[|\mathbf{P}(x)]; P(x)$]] $U^2_{N'} : N'$]]

This hypothesis raises several issues which we believe could be addressed by further experimental work. First of all, there is the question of the extent to which interpretation in a visual world context is the same as interpretation in other contexts, already raised by Britt (1994) (see (Tanenhaus and Trueswell 2005)). I.e., is a rule like $\mathbf{PARS1}_{Adj}$ only available in contexts in which a visual situation is available? Or perhaps only when the subject is required to do something with the objects in the situation?

We also hope that this example may explain why we believe that formulating the interpretation processes in more detail is going to show that current theories about incremental reference interpretation are still open. For instance, our formulation raises the question of whether there are in fact

separate versions of **PARS1**—i.e., different ways of using the visual situation for different types of expressions—or a single one. One may also wonder whether the principle proposed is only valid for intersective adjectives, or for all types, or even for all types of modifiers including for instance nominal premodifiers.[20]

6. Related Literature

We are not aware of any other account of the psychological results about incremental reference using the visual world paradigm in terms of a dynamic semantics, but there has been a lot of research relevant to providing such an account. In this section we will first of all discuss other work on incremental interpretation and formal grammar; then alternative theories of the dynamics of dialogues; finally, some recent computational models of the incremental interpretation of reference.

6.1 Related Linguistic Formalisms

Modulo the re-interpretation of trees in terms of MCEs, the grammar formalism used in PTT is (deliberately) very standard both in terms of LTAG analysis and in terms of CDRT semantics; in particular, it is closely related to Muskens' Logical Description Grammar (LDG) (Muskens 2001), an earlier proposal to combine LTAG with CDRT. The main differences concern the semantics of definites (Muskens' analysis is not based on Loebner's account or on resource situations). Also Muskens is not particularly concerned with incrementality; if at all, his formalism is more motivated by ideas about the role of underspecification in formal grammar.

The opposite is true of Dynamic Syntax (Cann et al. 2005), one of the few formal grammatical formalisms taking the incrementality of interpretation as the central fact that a theory of grammar has to explain.[21] The main difference between PTT and Dynamic Syntax lies in the treatment of anaphora. Contrary to what one could expect from the name, Dynamic Syntax is not based on a dynamic approach to the common ground in the sense of DRT or Dynamic Logic. Its concerns are mainly at the sentence level, and therefore, although it includes a proposal concerning the semantics of pronouns, it does not provide an account of which antecedents are available for them.[22]

6.2 Other Theories of the Common Ground in Dialogue

The two best-known theories of semantics in dialogue are SDRT (Asher and Lascarides 2003) and Ginzburg's KOS (Ginzburg 2011).

Like PTT, SDRT is an extension of DRT developed to account for the pragmatics of the common ground—in particular, for the effect of discourse structure on language interpretation. It thus provides a highly developed account of rhetorical relations and the process by which they get established, also based on the assumption that this is a process of defeasible inference. It does not however provide a theory of how interpretation proceeds incrementally, and it would not be easy to incorporate a treatment like PTTs, for although it would be quite simple to include the equivalent of micro conversational events in SDRT's picture of the common ground, one of the fundamental

20. One limitation of $\mathbf{PARS1}_{Adj}$ as formulated here is that it only applies to definite descriptions with a single adjective. This is not however a real limitation as it can easy be remedied by generalizing the rule to make it sensitive to the occurrence in the discourse situation of any NP containing an Adjectival Phrase (AdjP) rather than a single adjective.
21. Another being Combinatorial Categorial Grammar (Steedman 2001).
22. A more detailed comparison between PTT and Dynamic Syntax can be found in (Poesio and Rieser 2010).

assumptions of the theory is that the processes of grammatical interpretation and discourse interpretation are completely distinct—in fact, they are ruled by distinct logics.

KOS is built like PTT on the view of the common ground developed in Situation Semantics, and as such it incorporates very similar views about the presence of non-sentential utterances in the common ground (e.g., (Fernandez 2006)) but it is not built on a logical formalism designed to account for the anaphoric properties of utterances and until recently it did not incorporate an extensive treatment of anaphora.

A fairly detailed comparison between PTT and both these semantical formalisms can be found in (Poesio and Rieser 2010).

6.3 Computational Accounts of Incremental Reference

A computational implementation could eventually provide a large-scale test of the predictions of a model such as the one proposed in this paper. In recent years, the first computational models of incremental reference resolution have appeared. Although still relatively simple from a linguistic perspective, these models give us hope that computational modelling could soon become a tool in the study of incremental reference resolution.

Stoness et al. (2004) propose an account of the incremental interaction of reference resolution with parsing implemented in an actual spoken dialogue system. The account is based on the hypothesis that the reference resolution module is called upon every time that the parser identifies an NP, and attempts to find a referent for it in the knowledge base. The ability to resolve it adds to the score of that particular parsing interpretation, which may lead to it being chosen over the alternatives. Stoness *et al.* showed that this may result in improvements in parsing performance.

Schlangen et al. (2009) propose a model of incremental reference resolution based on a Bayesian Filtering model which captures quite directly the visual world scenario. Each object r in the visual scene is associated with a probability $P(r|w_{1:n})$ that words $w_1 \ldots w_n$ are referring to that object. This probability is incrementally updated after every word. Schlangen *et al.* proposed an evaluation metric for this task and methods for learning these probabilities.

Finally, Dubey (2010) implemented a computational model of reference interpretation consisting of a probabilistic parser, a probabilistic coreference resolver, and a pragmatics processor modelling coherence constraints. (The first two models are trained on actual data, the latter hand-coded.) He tested the model by simulating the garden path data finding a good match between predictions of the model and the experimental results.

6.4 Models of Visual Attention

A proper account of the effect of visual salience on the interpretation of referring expressions will require a more detailed theory of visual attention than the one assumed here. A proposal in such direction was made by Kelleher (Kelleher et al. 2005; Kelleher 2007).

7. Discussion

The proposal presented in this paper is, as far as we know, the only full account of the incremental interpretation of anaphoric and referential expressions taking into account the findings of both psycholinguistics and formal linguistics. The proposal makes a few clear predictions which should be possible to verify experimentally, including:

- that definite descriptions are not ambiguous between an anaphoric and referential interpretation;
- that it should be possible to categorize anaphors according to whether their resolution takes place at the surface level or at the deeper level.

The main limitation of the present proposal is that it does not provide a full list of the principles governing focus shift and resource situation anchoring, and that the model of defeasible reasoning adopted here is very simple. A natural development of the theory would be to provide an account couched in terms of a probabilistic model that could be learned from data (Jurafsky 1996; Bod et al. 2003).

A second limitation of the present work is that it is not integrated with an account of the grounding process such as that developed by Traum (1994), whose interaction with the present model of incremental interpretation was discussed in (Poesio and Rieser 2010). We think this development would be especially interesting at the light of the evidence from, e.g., Keysar et al. (2000) suggesting that reference does not only involve information in the common ground.

Finally, the current version of the theory also doesn't take full advantage of the formalization in terms of a defeasible logic to provide an account of reanalysis (Fodor and Ferreira 1998) and repairs (Ferreira et al. 2004). To do so however would require embedding the theory into a fully formalized account of belief revision.

Acknowledgments

We gratefully acknowledge the support of the CRC 673 *Alignment in Communication* at Universität Bielefeld, Germany, the University of Essex's School of Computer Science and Electronic Engineering, and the Center for Mind and Brain Sciences, Università di Trento, to the research leading to this paper. We also wish to thank David Schlangen, the editor of this paper, and the anonymous reviewers for their detailed comments and many insightful suggestions.

References

A. Abeille and O. Rambow, editors. *Tree Adjoining Grammars*. CSLI, 2000.

S. Abney. Parsing by chunks. In R. Berwick, S. Abney, and C.Tenny, editors, *Principle-based Parsing*, pages 257–278. Kluwer, Dordrecht, 1991.

J. F. Allen, L. K. Schubert, G. Ferguson, P. Heeman, C. H. Hwang, T. Kato, M. Light, N. Martin, B. Miller, M. Poesio, and D. R. Traum. The TRAINS project: a case study in building a conversational planning agent. *Journal of Experimental and Theoretical AI*, 7:7–48, 1995.

P. D. Allopenna, J. S. Magnuson, and M. K. Tanenhaus. Tracking the time course of spoken word recognition: evidence for continuous mapping models. *Journal of Memory and Language*, 38: 419–439, 1998.

G. T. M. Altmann and Y. Kamide. Incremental interpretation of verbs: Restricting the domain of subsequent reference. *Cognition*, 73:247–264, 1999.

G. T. M. Altmann and M. Steedman. Interaction with context during human sentence processing. *Cognition*, 30:191–238, 1988.

J. E. Arnold, J. G. Eisenband, S. Brown-Schmidt, and J. C. Trueswell. The immediate use of gender information: eyetracking evidence of the time-course of pronoun resolution. *Cognition*, 76:B13–B26, 2000.

N. Asher and A. Lascarides. *The Logic of Conversation*. Cambridge University Press, 2003.

J. L. Austin. *How to Do Things with Words*. Harvard University Press, Cambridge, MA, 1962.

J. Barwise. *The Situation in Logic*. CSLI Lecture Notes. University of Chicago Press, 1989.

J. Barwise and J. Perry. *Situations and Attitudes*. The MIT Press, 1983.

R. Beun and A. Cremers. Object reference in a shared domain of conversation. *Pragmatics and Cognition*, 6(1/2):121–152, 1998.

T. Bever. The cognitive basis for linguistic structure. In *Cognition and the Development of Language*. Wiley, New York, 1970.

R. Bod, J. Hay, and D. Jannedy, editors. *Probabilistic Linguistics*. MIT Press, 2003.

G. Brewka and T. Eiter. Prioritizing default logic. In S. Hölldobler, editor, *Intellectics and Computational Logic: Papers in Honor of Wolfgang Bibel*. Kluwer, 2000.

M. A. Britt. The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language*, 33:251–283, 1994.

S. Brown-Schmidt and M. Tanenhaus. Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science*, 32:643–684, 2008.

S. Brown-Schmidt, E. Campana, and M. K. Tanenhaus. Real-time reference resolution by naive participants. In J. C. Trueswell and M. K. Tanenhaus, editors, *Approaches to Studying World-Situated Language Use*, pages 153–171. MIT Press, 2005.

H. C. Bunt. Dialogue control functions and interaction design. In R.J. Beun, M. Baker, and M. Reiner, editors, *Dialogue in Instruction*, pages 197–214. Springer Verlag, 1995.

R. Cann, R. Kempson, and L. Marten. *The Dynamics of Language: an Introduction*. Elsevier, 2005.

B. Carpenter. *Type-Logical Semantics*. MIT Press, 1998. URL `http://mitpress.mit.edu/book-home.tcl?isbn=0262531496`.

C. G. Chambers, M. K. Tanenhaus, K. M. Eberhard, H. Filip, and G. N. Carlson. Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language*, 47:30–49, 2002.

G. Chierchia. *Dynamics of meaning. Anaphora, Presupposition and the Theory of Grammar*. University of Chicago Press, 1995.

H. H. Clark. *Using Language*. Cambridge University Press, Cambridge, 1996.

H. H. Clark and C. R. Marshall. Definite reference and mutual knowledge. In A. Joshi, B. Webber, and I. Sag, editors, *Elements of Discourse Understanding*. Cambridge University Press, New York, 1981.

R. Cooper. The role of situations in generalized quantifiers. In S. Lappin, editor, *Handbook of Contemporary Semantic Theory*, chapter 3, pages 65–86. Blackwell, 1996.

A. Corbett and F. Chang. Pronoun disambiguating: Accessing potential antecedents. *Memory and Cognition*, 11:283–294, 1983.

S. Crain and M. Steedman. On not being led up the garden path: the use of context by the psychological syntax processor. In D. R. Dowty, L. Karttunen, and A. M. Zwicky, editors, *Natural Language Parsing: Psychological, Computational and Theoretical perspectives*, pages 320–358. Cambridge University Press, New York, 1985.

A. Dubey. The influence of discourse on syntax: a psycholinguistic model of sentence processing. In *Proc. of the ACL*, Uppsala, Sweden, 2010.

K. Eberhard, S. Spivey-Knowlton, J. Sedivy, and M. Tanenhaus. Eye movements as a window into real-time spoken language processing in natural contexts. *Journal of Psycholinguistic Research*, 24:409–436, 1995.

R. Fernandez. *Non-Sentential Utterances in Dialogue: Classification, Resolution and Use*. PhD thesis, Department of Computer Science, King's College, London, 2006.

F. Ferreira, E. F. Lau, and K. G. D. Bailey. Disfluencies, language comprehension, and tree adjoining grammars. *Cognitive Science*, 28:721–749, 2004.

J. D. Fodor and F. Ferreira, editors. *Reanalysis in Sentence Processing*. Kluwer, 1998.

L. Frazier. *On comprehending sentences: syntactic parsing strategies*. PhD thesis, University of Connecticut, 1979. Available via Indiana University Linguistic Club.

L. Frazier. Sentence processing: a tutorial review. In M. Coltheart, editor, *Attention and performance XII: The Psychology of Reading*, pages 559–586. Erlbaum, Hove, 1987.

L. Frazier and K. Rayner. Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language*, 29:181–200, 1990.

S. C. Garrod. Resolving pronouns and other anaphoric devices: The case for diversity in discourse processing. In C. Clifton, L. Frazier, and K. Rayner, editors, *Perspectives in Sentence Processing*. Lawrence Erlbaum, 1994.

S.C. Garrod and A. J. Sanford. Resolving sentences in a discourse context. In M. A. Gernsbacher, editor, *Handbook of Psycholinguistics*, chapter 20, pages 675–698. Academic Press, 1994.

J. M. Gawron and S. Peters. *Anaphora and Quantification in Situation Semantics*, volume 19 of *Lecture Notes*. CSLI, 1990.

E. Gibson. *A computational theory of human linguistic processing: memory limitations and processing breakdown*. PhD thesis, Carnegie Mellon University, Pittsburgh, 1991.

J. Ginzburg. *The Interactive Stance: Meaning for Conversation*. Oxford, 2011.

J. Ginzburg. Situation semantics: from indexicality to metacommunicative interaction. In P. Portner, C. Maierborn, and K. von Heusinger, editors, *The Handbook of Semantics*. de Gruyter, To Appear. To Appear.

D. Gross, J. Allen, and D. Traum. The TRAINS 91 dialogues. TRAINS Technical Note 92-1, Computer Science Dept. University of Rochester, June 1993.

B. J. Grosz. *The Representation and Use of Focus in Dialogue Understanding*. PhD thesis, Stanford University, 1977.

B. J. Grosz and C. L. Sidner. Attention, intention, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, 1986.

B. J. Grosz, A. K. Joshi, and S. Weinstein. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):202–225, 1995. (The paper originally appeared as an unpublished manuscript in 1986.).

J. K. Gundel, N. Hedberg, and R. Zacharski. Cognitive status and the form of referring expressions in discourse. *Language*, 69(2):274–307, 1993.

Jorge Hankamer and Ivan Sag. Deep and surface anaphora. *Linguistic Inquiry*, 7(3):391–426, 1976.

J. R. Hobbs, M. Stickel, P. Martin, and D. Edwards. Interpretation as abduction. *Artificial Intelligence Journal*, 63:69–142, 1993.

W. S. Horton and B. Keysar. When do speakers take into account common ground? *Cognition*, 59: 91–117, 1996.

J. F. Horty. Defaults with priorities. *Journal of Philosophical Logic*, 36:367–413, 2007.

V. Hoste. *Optimization Issues in Machine Learning of Coreference*. PhD thesis, University of Antwerp, 2005.

C. H. Hwang and L. K. Schubert. Episodic logic: A comprehensive, natural representation for language understanding. *Minds and Machines*, 3:381–419, 1993.

D. Jurafsky. A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, 20:137–194, 1996.

D. Jurafsky and J. H. Martin. *Speech and Language Processing*. Prentice Hall, 2nd edition, 2009.

H. Kamp and U. Reyle. *From Discourse to Logic*. D. Reidel, Dordrecht, 1993.

D. Kaplan. On the logic of demonstratives. *Journal of Philosophical Logic*, 8:81–98, 1978.

J. D. Kelleher. Attention driven reference resolution in multimodal contexts. *Artificial Intelligence Review*, 25:21–35, 2007.

J.D. Kelleher, F. Costello, and J. van Genabith. Dynamically updating and interrelating representations of visual and linguistic discourse. *Artificial Intelligence*, 167:62–102, 2005.

B. Keysar, D. J. Barr, J. A. Balin, and J. S. Brauner. Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, 11:32–37, 2000.

H. E. Kyburg. *Logical Foundations of Statistical Inference*. D. Reidel, 1974.

S. Larsson and D. R. Traum. Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural Language Engineering*, 6:323–340, 2000.

A. Lascarides and N. Asher. Discourse relations and defeasible knowledge. In *Proc. ACL-91*, pages 55–63, University of California at Berkeley, 1991.

S. Loebner. Definites. *Journal of Semantics*, 4:279–326, 1987.

A. Lücking, K. Bergmann, F. Hahn, S. Kopp, and H. Rieser. The bielefeld speech and gesture alignment corpus (saga). In *Proc. of the LREC Workshop on Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*, 2010.

A. Lücking, T. Pfeiffer, and H. Rieser. Pointing and reference reconsidered. Submitted, To Appear.

M. C. MacDonald, N. J. Pearlmutter, and M. S. Seidenberg. Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4):676–703, 1994.

W.D. Marslen-Wilson. Sentence perception as an interactive parallel process. *Science*, 189:226–228, 1975.

W.D. Marslen-Wilson. Functional parallelism in spoken word recognition. *Cognition*, 25:71–102, 1987.

C. Matheson, M. Poesio, and D. Traum. Modeling grounding and discourse obligations using update rules. In *Proc. of the First Annual Meeting of the North American Chapter of the ACL*, Seattle, April 2000.

J. L. McClelland and J. L. Elman. Interactive processes in speech perception: the trace model. In D. E. Rumelhart and J. L. McClelland, editors, *Parallel Distributed Processing*, volume 2. MIT Press, 1986.

D. Milward. Dynamic Dependency Grammar. *Linguistics and Philosophy*, 17:561–605, 1994.

R. A. Muskens. Combining Montague Semantics and Discourse Representation. *Linguistics and Philosophy*, 19:143–186, 1996.

R. A. Muskens. Talking about trees and truth conditions. *Journal of Logic, Language and Information*, 10(4):417–455, 2001.

G. D. Nunberg. Transfers of meaning. *Journal of Semantics*, 12(2):109–132, 1995.

B. H. Partee. Topic, focus and quantification. In *Proc. SALT-91*, 1991.

R. J. Passonneau. Getting and keeping the center of attention. In M. Bates and R. M. Weischedel, editors, *Challenges in Natural Language Processing*, chapter 7, pages 179–227. Cambridge University Press, 1993.

N. J. Pearlmutter and A. A. Mendelsohn. Serial versus parallel sentence comprehension. Manuscript in revision, available at `http://www.psych.neu.edu/People/njp/papers`, 1999.

F. C. N. Pereira. Categorial semantics and scoping. *Computational Linguistics*, 16(1):1–10, March 1990.

C. R. Perrault. An application of default logic to speech act theory. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, chapter 9, pages 161–185. The MIT Press, Cambridge, MA, 1990.

T. Pfeiffer. *Understanding Multimodal Deixis with Gaze and Gesture in Conversational Interfaces*. PhD thesis, University of Bielefeld, 2010.

M. Poesio. A situation-theoretic formalization of definite description interpretation in plan elaboration dialogues. In P. Aczel, D. Israel, Y. Katagiri, and S. Peters, editors, *Situation Theory and its Applications, vol.3*, chapter 12, pages 339–374. CSLI, Stanford, 1993.

M. Poesio. *Discourse Interpretation and the Scope of Operators*. PhD thesis, University of Rochester, Department of Computer Science, Rochester, NY, 1994.

M. Poesio. A model of conversation processing based on micro conversational events. In *Proceedings of the 17th Annual Conference of the Cognitive Science Society*, pages 698–703, Pittsburgh, July 1995a.

M. Poesio. Disambiguation as (defeasible) reasoning about underspecified representations. In P. Dekker and M.Stokhof, editors, *Proc. of the Tenth Amsterdam Colloquium*, pages 607–625. ILLC, December 1995b.

M. Poesio. Semantic ambiguity and perceived ambiguity. In K. van Deemter and S. Peters, editors, *Semantic Ambiguity and Underspecification*, chapter 8, pages 159–201. CSLI, Stanford, CA, 1996.

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

M. Poesio and R. Artstein. The reliability of anaphoric annotation, reconsidered: Taking ambiguity into account. In A. Meyers, editor, *Proc. of ACL Workshop on Frontiers in Corpus Annotation*, pages 76–83, June 2005.

M. Poesio and M. A. Kabadjov. A general-purpose, off the shelf anaphoric resolver. In *Proc. of LREC*, pages 653–656, Lisbon, May 2004.

M. Poesio and R. Muskens. The dynamics of discourse situations. In P. Dekker and M. Stokhof, editors, *Proceedings of the 11th Amsterdam Colloquium*, pages 247–252. University of Amsterdam, ILLC, December 1997.

M. Poesio and H. Rieser. Anaphora and direct reference: Empirical evidence from pointing. In *Proc. of DiaHolmia, the 13th Workshop on the Semantics and Pragmatics of Dialogue*, pages 35–43, Stockholm, June 2009.

M. Poesio and H. Rieser. Completions, coordination, and alignment in dialogue. *Dialogue and Discourse*, 1(1):1–89, 2010. doi: 10.5087/dad.2010.001.

M. Poesio and D. Traum. Conversational actions and discourse situations. *Computational Intelligence*, 13(3):309–347, 1997.

M. Poesio, R. Stevenson, B. Di Eugenio, and J. M. Hitzeman. Centering: A parametric theory and its instantiations. *Computational Linguistics*, 30(3):309–363, 2004.

M. Poesio, P. Sturt, R. Arstein, and R. Filik. Underspecification and anaphora: Theoretical issues and preliminary evidence. *Discourse Processes*, 42(2):157–175, 2006.

M. E. Pollack. *Inferring Domain Plans in Question-Answering*. PhD thesis, Department of Computer and Information Science, University of Pennsylvania, 1986.

J. L. Pollock. *How to Build a Person*. MIT Press, 1990.

R. Reichman. *Getting Computers to Talk Like You and Me*. The MIT Press, Cambridge, MA, 1985.

M. Richardson and P. Domingos. Markov logic networks. *Machine Learning*, 62:107–136, 2006.

C. Roberts. Demonstratives as definites. In K. van Deemter and R. Kibble, editors, *Information Sharing*, pages 89–196. CSLI, 2002.

M. Rooth. A theory of focus interpretation. *Natural Language Semantics*, 1:75–116, 1992.

Y. Schabes. New parsing strategies for Tree Adjoining Grammars. In *Proc. of the 12th International Conference on Computational Linguistics (COLING)*, 1988.

D. Schlangen, T. Baumann, and M. Atterer. Incremental reference resolution: The task, metrics for evaluation, and a bayesian filtering model that is sensitive to disfluencies. In *Proc. of SIGDIAL*, London, UK, 2009.

M. S. Seidenberg, M. K. Tanenhaus, J. Leiman, and M. Bienkowski. Automatic access of the meanings of ambiguous words in context: some limitations of knowledge-based processing. *Cognitive Psychology*, 14:489–537, 1982.

S. Shieber and M. Johnson. Variations on incremental interpretation. *Journal of Psycholinguistic Research*, 22(2):287–318, 1993.

C. L. Sidner. *Towards a computational theory of definite anaphora comprehension in English discourse*. PhD thesis, MIT, 1979.

M. J. Spivey, M. K. Tanenhaus, K. M. Eberhard, and J. C. Sedivy. Eye movements and spoken language comprehension: Effect of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, 45:447–481, 2002.

M. Steedman. *The Syntactic Process*. MIT Press, 2001.

M. Stone. Intention, interpretation and the computational structure of language. *Cognitive Science*, 28(5):781–809, 2004.

S. C. Stoness, J. Tetreault, and J. Allen. Incremental parsing with reference interaction. In *Proc. ACL Workshop on Incremental Parsing*, pages 18–25, Barcelona, 2004.

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

D. A. Swinney. Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18:545–567, 1979.

M. K. Tanenhaus and J. C. Trueswell. Eye movements as a tool for bridging the language-as-product and language-as-action traditions. In J. C. Trueswell and M. K. Tanenhaus, editors, *Approaches to Studying Wordl-Situated Language Use*, pages 3–37. MIT Press, 2005.

M. K. Tanenhaus, M. Spivey-Knowlton, K. M. Eberhard, and J. C. Sedivy. Integration of visual and linguistic information in spoken language comprehension. *Science*, 268:1632–1634, June 1995.

M. K. Tanenhaus, C. G. Chambers, and J. E. Hanna. Referential domains in spoken language comprehension: Using eye movements to bridge the product and action traditions. In J. M. Henderson and F. Ferreira, editors, *The Interface of language, vision, and action: Eye movements and the visual world*, pages 279–317. Psychology Press, 2004.

D. R. Traum. *A Computational Theory of Grounding in Natural Language Conversation*. PhD thesis, University of Rochester, Department of Computer Science, Rochester, NY, July 1994.

D. R. Traum and E. A. Hinkelman. Conversation acts in task-oriented spoken dialogue. *Computational Intelligence*, 8(3), 1992. Special Issue on Non-literal Language.

B. L. Webber. *A Formal Approach to Discourse Anaphora*. Garland, New York, 1979.

B. L. Webber. Structure and ostension in the interpretation of discourse deixis. *Language and Cognitive Processes*, 6(2):107–135, 1991.

H. Zender. *Situated Production and Understanding of Verbal References to Entities in Large-Scale Space*. PhD thesis, University of Saarbruecken, 2010.