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

Massimo Poesio
Università di Trento

Hannes Rieser
Universität Bielefeld

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), 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
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 subjects nevertheless find odd because the ambiguity between a reduced relative reading and a matrix verb reading of the verbs \textit{raced}, \textit{floated} etc. is immediately resolved in favor of the matrix verb interpretation, thus forcing the reader to a reanalysis step later on.

(2) a. \textit{The horse raced past the barn fell.}  
    b. \textit{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 \textit{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 \textit{spiders, roaches and other}). Swinney found priming effects for both \textit{ant} and \textit{spy} at [1], even with a strongly disambiguated context; but only for \textit{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 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).


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).

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*.

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 (27).

(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 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 experimen-
tal situations. The experiments discussed in (Brown-Schmidt et al. 2005; Brown-Schmidt and Tanenhaus 2008), by contrast, involved subjects performing tasks in fairly ecological 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). 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 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 conversation 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 (8) (an
edited version of two turns from the Bielefeld ToyPlane Corpus) result in
the update to the common ground in (9).

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
and you put it through from above so that the three get fixed

drugs specifying the interpretation of the two utterances in (8).

The contents of conversational events are associated with propositional dis-
course referents (discourse referents whose values are

It is further assumed in PTT that dialogue acts are
generated (Pollack 1986) by locutionary acts (Austin 1962) which we repre-
sent 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:

where $\alpha$ is what DG is pointing at–determining experimentally what is
$\alpha$ being the main question addressed by (Lücking et al. To Appear), as we
will see.

8
3.2 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 1995b). 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 1995b)) 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. One type of syntactic information about MCEs is syntactic constituency: every MCE $u$ that is not the root of a tree has a mother node $u'$. We indicate this following the notation from (Muskens 2001):

\[ u \uparrow u' \]

We will assume here the following additional features of MCEs:

- **cat** specifying the syntactic category of a MCE;
- **gen** specifying the gender of MCEs;
- **num** specifying the number;
- **sem** specifying its (conventional) semantics; and
- **do** specifying the discourse referent introduced by the NP, if any.

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).

---

1 There is a clear relation between MCEs in PTT and signs in theories such as HPSG, see (Poesio To Appear) for discussion.
We will mostly represent the information associated with mces in the compact representation illustrated by the following example, representing the update resulting from observing an utterance of determiner *the* and by the following lexical access.

\[
\begin{array}{c}
\text{the:}\lambda P\lambda P'[y|y \text{ is } \lambda x P(x)]; P'(y)
\\
\text{u}_{\text{Det}}:\text{Det}
\\
\text{u}_{\text{NP}}:\text{NP}
\\
\text{u}_{\text{N}}:\text{N'}
\\
\text{u}_{\text{NP}}:\text{NP}
\end{array}
\]

3.3 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, 1995b,a, 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. 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,\(^2\) 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, 1995a; 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

\(^2\)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).
that utterance, where each update adds to the discourse situation the information exemplified by (11)—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.  

$$U : \text{utter}(A, \text{“the”}) : [U, U_{NP}, U_{N’}] \quad U_{NP} : NP$$

$$\begin{array}{c}
U : \text{Det} \\
\text{the} \\
\end{array}$$

$$U_{N’} : N’$$

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:

$$\begin{array}{c}
U : \text{utter}(A, \text{“the”}) : [U, u_{NP}, u_{N’}] \\
\end{array}$$

$$U_{NP} : NP$$

$$\begin{array}{c}
U : \text{Det} \\
\text{the} \\
\end{array}$$

$$u_{N’} : N’$$

The PTT view of the interpretive processes that follow lexical access such as syntactic interpretation (parsing) is very much inspired by current 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

---

3Most 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.
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 (12), to be imagined uttered by a customer handing his key to an attendant at a parking lot:

(12) 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.

(13) λy.(∀x[car-of(y) = x]→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 ...

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 (13)).

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 $\text{sem}(U_3)$ of $U_3$ is that it results from the application of the semantic value of $U_1$ to the semantic value of $U_2$.

\[
\begin{align*}
U_1 \uparrow U_3, & \text{ } U_2 \uparrow U_3, \text{ } \text{sem}(U_1) \text{ is } \phi_{(\alpha,\beta)}, \text{ } \text{sem}(U_2) \text{ is } \psi_\alpha : \text{ } \text{sem}(U_3) \text{ is } \phi(\psi)_{\text{BSC}} \\
\text{sem}(U_3) \text{ is } \phi(\psi)
\end{align*}
\]

(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).\(^4\)

\[
\begin{align*}
U_1 \uparrow U_2, & \text{ } \text{sem}(U_1) \text{ is } \phi_\alpha : \text{ } \text{sem}(U_2) \text{ is } \phi_{\text{USC}} \\
\text{sem}(U_2) \text{ is } \phi
\end{align*}
\]

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.

3.4 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)).\(^5\)

\(^4\)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).

\(^5\)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).
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 \textbf{h-underspecified} in the sense of (Poesio To Appear)—i.e., the discourse situation does not specify the value of \( \text{sem}(u) \), 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 higher 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 processor \textbf{commits} itself to that hypothesis, 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 \textbf{p-underspecified} interpretation (Poesio To Appear; Poesio \textit{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.  

\footnote{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.}
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 novelty 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,⁷ the idea is that proper noun Jack denotes the (0-argument) function

\( \lambda x. (x = j) \),

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

\( \lambda s. \lambda x. (x = \text{pope}(s)(x)) \),

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

\[ \text{Jack} \Rightarrow \lambda P. ([y | y \text{ is } \lambda x. (x = j)]; P(y)) \]

whereas definite descriptions like the chair translate as follows:

⁷Loebner only provided an informal discussion of his theory.
the chair ⇒ λP.(y|y is [x].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 (15), which combines an ltag-style prediction of an elementary tree with the cdrt semantics just discussed. The update to the discourse situation in (15) 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 $\lambda P\lambda P'(y|y is [x].P(x)); P'(y))$.

(15) \[
\begin{array}{c}
\langle u_{Det}, u_{NP}, u_N \rangle \\
\hline
u_{NP}:NP \\
\end{array}
\]
\[
\begin{array}{c}
\langle u_{Det}:Det \quad u_{NP}:N' \rangle \\
\hline
\end{array}
\]
\[
\text{the:} \lambda P\lambda P'(y|y is [x].P(x)); P'(y)
\]

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 (16). We’ll discuss pronouns in Section 4.5.

(16) \[
\begin{array}{c}
\langle u_{PN}, u_{NP} \rangle \\
\hline
u_{NP}:NP \\
\end{array}
\]
\[
\begin{array}{c}
\langle \quad \rangle \\
\hline
u_{PN}:PN \\
\end{array}
\]
\[
\text{Jack: } \lambda P(y|y is [x].is j); P(y)
\]

### 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 ambiguous, but depend for their interpretation 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 situation s described by core speech act (csa, (Traum and Hinkelmann 1992)) c as a possible resolution for a resource situation.

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 if the NP is lexically primed by 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 (17).

\[
(17) \quad \text{the chair} \Rightarrow \lambda P'. ([y|y \text{ is } \alpha x.K; \text{chair}(x)]; P'(y))
\]

According to the semantics in (17), 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 default ones discussed earlier, BSC and USC, and that coerce a nominal predicate $P$ into a presuppositional interpretation $\lambda x. K; [[P(x)]]$ that is pragmatically functional wrt a resource situation $K$, turning the NP interpretation in (15) into the following one:

$$(18)\quad u_{NP}:NP$$

$$u_{Det}:Det\ |\ u_{NP}:N : \lambda xK; [[P(x)]];\ |\ x is Z$$

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 $P$ into interpretation

$$\lambda xK_{mva} ; [[P(x)]]; [[x is Z]]$$

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 $U_{NP}$ 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 unnecessary. We also use the notation $K \models \phi$, for $K$ a dRS and $\phi$ a condition, to indicate that

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

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:

$${\text{MSOA is } K_{mva} \rightarrow Bel_{A,B}(A,B, see_{A,B}(A,B,K))}$$

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

\[
[U_{NP} : NP \ [\text{Det the: } \lambda P \lambda P'(\{y | y \text{ is } \iota x P(x)\}; P'(y))] \ [U_{N'} : N' [N : \lambda x [[P(x)]]]] \\
MSO \text{ is } K_{mva} \\
K_{mva} \models P(Z) \\
\Rightarrow \\
[U_{NP} : NP \ [\text{Det the: } \lambda P \lambda P'(\{y | y \text{ is } \iota x P(x)\}; P'(y))] \\
[U_{N'} : N' : \lambda x K_{mva}; [P(x)]; [x \text{ is } Z] [N : \lambda x [[P(x)]]]]
\]

Notice that the uniqueness requirement on \(Z\), proposed in (Poesio 1994), has been dropped. This is because any hypothesis concerning a situation in which there is more than one object of type \(P\) would coerce \(P\) in a non-functional predicate which would not satisfy the requirements of the \(\iota\) operator and would therefore be filtered out semantically. This appears to be consistent with the results of eye-tracking experiments in which multiple objects in the visual situation are briefly considered (Tanenhaus and Trueswell 2005).

Notice also that the formulation of PARS1 above would only work, strictly speaking, for definite descriptions. This is in keeping with the assumption that distinct interpretation processes apply to each 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 (8), 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 (9), 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, $ce2.1$).

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 to interpret, $U_{NP}$, 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

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

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

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

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 $U_{NP}$.
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.

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 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:

\[ \text{INF (INF)} \text{ An object is referred to by pointing only if} \]

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.

\[ \text{INF succeeds in 96\% of the cases, which led Poesio and Rieser to formulate what they called Strong prag hypothesis:} \]
**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 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 in 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.)

$$\text{PARS3}$$

\[
[U_{NP} : NP \ [\text{Det this: } \lambda P \lambda P' (\{y | y \text{ is } \iota x P(x); P'(y)\})]]
\]

\[
[U_{N'} : N' : \lambda x K_{pointing}; [\{P(x); [\{x \text{ is } Z\} \{N : \lambda x [P(x)]\}]\}]]
\]

Notice that as formulated **PARS3** only applies to demonstratives with *this*. We believe a version of the default may exist for demonstratives with *this*, 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 (19) makes pronouns behave semantically like determiners, as well.

(19) $u_{NP}:NP$

\[
[u_{Pr} : Pro]
\]

\[
sie; \lambda P \lambda P' ([y | y \text{ is } \iota x K; P(x); P'(y)])
\]

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

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

for $z$ a discourse entity. According to the treatment just sketched, resolving *sie* in (8) involves identifying the content of the first directive in (9), $K1.1$, as the resource situation for the pronoun, and discourse entity $z$ as the antecedent (i.e., applying the result to the identity property $\lambda w([w \text{ is } z])$).

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 $\textit{PARS2}$, but which applies specifically to pronouns. We call the principle $\textit{PARS2}_\text{pro}$. In first approximation, here is a version of the principle that is exactly as $\textit{PARS2}$. This version of the principle generates a semantic interpretation for any pronominal form $\textit{PRO}$ which is part of the performance of an utterance $U$ generating core speech act $C'$ such that an antecedent for $\textit{PRO}$ is available as part of the content of speech act $C$ accessible from $C'$.

$$\textit{PARS2}_\text{pro}$$

$$[U NP : NP [\text{Pro PRO: } \lambda P \lambda P'(y; y \text{ is } \iota x; P(x)); P'(y)])]$$

$U NP \uparrow U$, generates($U, C'$), $C : \textit{csa}(A, B, K_{at}), \textit{accessible}(C, C'), K_{at} = P(Z)$

$$\Rightarrow [U NP : NP : \lambda P'[y; y \text{ is } \iota x; K_{1.1}; [x \text{ is } z]; P'(y))]

[\text{Pro PRO: } \lambda P \lambda P'(y; y \text{ is } \iota x; P(x)); P'(y))]

$$

This principle produces the interpretation in (19).

$$(19') u_{NP}:NP: \lambda P'[y; y \text{ is } \iota x; K_{1.1}; [x \text{ is } z]; P'(y))$$

$$u_{Det}:Det$$

$$\textit{sie}: \textit{PRO: } \lambda P \lambda P'(y; y \text{ is } \iota x; P(x)); P'(y))$$

We will propose a revised formulation of this Principle, including also an agreement check, in Section 5.4, after presenting our view for 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 (20).

\[ u_{the} : \text{utter}(A, "the") \]

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 3.3. This in turn leads to the update in (15) 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.)

\[ u_{Det}, u_{NP}, u_{N'} \]

\[ \text{Det} \]

\[ \text{NP} \]

\[ \text{N'} \]

\[ \text{the} : \lambda P \lambda P'[y | y \text{ is } \exists x P(x)] : P'(y) \]
Next (or while the interpretation inferences activated by the are taking place), perceiving the utterance of noun apple leads to the update in (21), 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 (22) through wordsense disambiguation processes which are not our concern here.\(^8\)

\[
\text{(21)} \quad \boxed{u_{\text{apple}} : \text{utter}(A, \text{“apple”})}
\]

\[
\text{(22)} \quad \boxed{u_N \quad u_N : N \quad \text{apple} : \lambda x[[\text{apple}(x)]]}
\]

Parsing then results in the interpretation in (23), through one of the basic LTAG operations, substitution of \(u_N\) into \(u_{NP}\).

\[
\text{(23)} \quad \boxed{u_{Det}, u_{NP}, u_{NP'}} \quad u_{NP}\text{:NP} \quad u_{Det}\text{:Det} \quad u_{NP'}\text{:N'} \quad \text{the} : \lambda P\lambda P'[y][y \text{ is } \lambda xP(x)]; P'(y) \quad u_{N'}\text{:N'} \quad u_N : N \quad \text{apple} : \lambda x[[\text{apple}(x)]]}
\]

In both visual world scenarios in Figure 1 the entire visual world–called here \(K_{\text{visual}}\)–is in the mutual focus of attention

\[
MSOA = K_{\text{visual}}
\]

so that the subject can use PARS1 to assign \(K_{\text{visual}}\) 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 (24). This interpretation is fully specified: the discourse situation provides syntactic and semantic interpretations for all phrasal and lexical MCES. The subject can therefore commit to the interpretation in (24), 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.)

\(^8\)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).
In (24), 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 on the right, \textsc{pars1} can be applied twice to produce two hypotheses concerning the \text{sem} value of utterance \( u_N' \)–as in (24), and the interpretation which is identical to the one (24) 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. However neither of these interpretations satisfies the uniqueness requirement on the interpretation of the definite: neither \( a_1 \) nor \( a_2 \) are the only apple in situation \( K_{visual} \). As a result, subjects cannot assign an interpretation to the NP utterance \( u_{NP} \), and therefore have to backtrack, so that when the rest of the instruction, ... on the towel in the box, comes in, hypothesis (23) 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 \textsc{trains} dialogues appeared to be tied in with locations on the map–i.e., moving a train to a location seemed to restrict the domain of interpretation to the area of the map around that location. This led to the formulation of the hypothesis that the \textsc{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 \textsl{MSOA}; it was also possible that larger sub-situations could be identified. And whereas Grosz (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 \textsc{follow-the-movement}:

\begin{center}
\textsc{follow-the-movement} Part of the intended effect of an utter-
\end{center}
 ance 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 closests 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 by Cavicchio et al. (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

\[ \text{Int}_A(move(\{A,B\},C,L)) \Rightarrow \text{Int}_A(\text{MSOA is } K_L) \]

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:

\[ K_{fit-cube} \mid K_{fit-cube} \text{ is } [x, y, s \mid x \text{ is } \tau.z.p\text{ick-up(subj, } z), \text{container}(y), s : \text{fit-in}(x, y)] \]

(25)
and this new resource situation then becomes the MSOA, ensuring that PARS2\textsubscript{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 \textit{et al.} 1995; Spivey \textit{et al.} 2002). Let us consider again the two types of experimental settings contrasted 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 subject can use PARS1 to choose the current MSOA, \(K\text{\_visual} \), as the resource situation, and choose the only apple in \(K\text{\_visual} \), that we called \(a_1\), to produce the single hypothesis in (24), repeated below for convenience.

\[
\begin{align*}
\text{app}_{NP} : & \text{Det}_{NP} : \lambda P'[y] \mid [y \text{ is } \iota x K\text{\_visual} \mid \text{apple}(x) \mid [x \text{ is } a_1] ; P'(y) \\
\text{app}_{Det} : & \text{Det} \; u_{NP} : \lambda x K\text{\_visual} \mid \text{apple}(x) \mid [x \text{ is } a_1] ; P'(y) \\
\text{app}_{NP} : & \lambda x \; \text{apple}(x) \\
\end{align*}
\]

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 (24), preventing further attachments. So when the subject hears next the utterance of a PP, \textit{on the towel}, the only available interpretation is as an argument of \textit{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 ((24) and the interpretation which is identical to (24) in all respects except that apple \(a_2\) is chosen) could not be accepted as they did not satisfy the uniqueness restriction imposed by the determiner, and therefore subjects have to backtrack to hypothesis (23). This means that when the next part of the instruction, \ldots\textit{on the towel}, comes in, this hypothesis is still open to modification—in fact, it requires the meaning of \(u_{NP}\) to be restricted in order to find a discourse referent satisfying the uniqueness condition. We argue that this requirement is what makes adjunction of \ldots\textit{on ...} to \(u_{NP}\) preferred over substitution as second argument of \textit{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 on, as follows.

\[
(26) \quad [K_{\text{apple-on}}, \text{MSOA}]
\]

\[
K_{\text{apple-on}} \text{ is } [x, y, s | x \text{ is } \zeta z. [\text{apple}(z), \text{put}(\text{subj}, z, y)], \text{object}(y), s : \text{on}(x, y)],
\]

\[
\text{MSOA is } K_{\text{apple-on}}
\]

Notice that the new resource situation only contains one towel, the towel in the top left quadrant, that we will call \(t_1\). \(K_{\text{apple-on}}\) can now be chosen as resource situation for *the towel* via **PARS1**: this makes the definite felicitous.\(^9\)

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*.

\[
(26) \quad [K_{\text{apple-on}}, \text{MSOA}]
\]

\[
K_{\text{apple-on}} \text{ is } [x, y, s | x \text{ is } \zeta z. [\text{apple}(z), \text{put}(\text{subj}, z, y)], \text{object}(y), s : \text{on}(x, y)],
\]

\[
\text{MSOA is } K_{\text{apple-on}}
\]

\[^9\text{The idea that } \text{the towel} \text{ in this case is felicitous in virtue of being interpreted as, essentially, 'the towel that an apple is on' is reminiscent of the interpretation for definites proposed by Webber in her thesis (Webber 1979).}\]
As this update makes the meaning of $u_N'$ functional, **PARS1** can now apply to choose the visual situation $K_{\text{visual}}$ 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.

We now discuss how the new version of **PARS2** for pronouns proposed in Section 4.5, **PARS2**<sub>pro</sub>, can explain how an interpretation is assigned to the pronouns in the experiments discussed by Arnold *et al.*, repeated here.

(27) 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, that shares many assumptions with PTT.

The first utterance, (27a), generates a core speech act of type **assert**, that we will call $ce_1$. Overall, the update resulting from the first utterance is then as in (28a), 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 (27b), is the same as (27a) except that in this case x3 refers to Minnie and the gender of up1.2 is feminine.

(28) a. \[ K_1, up1, ce1, up1.1, up1.2 \]
    \( K_1 \) is \( [e_1, x_1, x_2, x_3|x_1 \text{ is } Donald, \text{mail}(x_2), x_3 \text{ is } Mickey, \]
    \( e_1 : \text{bring}(x_1, x_2, x_3)] \),
    up1.1 : \text{utter}(A, "Donald"),
    gen(up1.1) is masc,
    up1.1 ↑ up1,
    up1.2 : \text{utter}(A, "Mickey"),
    gen(up1.2) is masc,
    up1.2 ↑ up1,
    up1 : \text{utter}(A, "Donald is bringing some mail to Mickey"),
    sem(up1) is K1,
    ce1 : assert(A, B, K1)
    generate(up1, ce1) \]

b. \[ K_2, up1, ce1, up1.1, up1.2 \]
    \( K_2 \) is \( [e_1, x_1, x_2, x_3|x_1 \text{ is } Donald, \text{mail}(x_2), x_3 \text{ is } Minnie \]
    \( e_1 : \text{bring}(x_1, x_2, x_3)] \),
    up1.1 : \text{utter}(A, "Donald"),
    gen(up1.1) is masc,
    up1.1 ↑ up1,
    up1.2 : \text{utter}(A, "Minnie"),
    gen(up1.2) is fem,
    up1.2 ↑ up1,
    up1 : \text{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. We show the update resulting from He in (29a), that resulting from She in (29b).

(29) a. \[ K_2, up2, ce2 \] utterance(up2), ce2 : csa(A, B, K2),
    generate(up2, ce2), up2.1 : \text{utter}(A, "He"), up2.1 ↑ up2,
    gen(up2.1) is masc \]
b. \[ K_2, up2, ce2 \] utterance(up2), ce2 : csa(A, B, K2),
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 identified 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 end this does matter in the case of anaphoric references, like pronoun *sie* in (8) or the pronouns in the example under discussion. This is because principles \textbf{PARS2} and \textbf{PARS2}$_{pro}$, in order to choose the content of a previous core speech act $C$ as resource situation, require that core speech act to \textbf{dominate} or \textbf{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 \textsc{sdt}, and consistently with the general view of interpretation discussed in Section 3, discourse structure recognition is viewed in \textsc{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 \textsc{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 \textsc{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 \textsc{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.)

The treatment of ‘surface’ anaphora resolution (Hankamer and Sag 1976) we propose here is based, on the one end, on the proposal by Garrod (1994) and Garrod and Sanford (1994) that the resolution of these types of anaphoric reference consists of distinct \textbf{bonding} and \textbf{resolution} stages;
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.\(^{10}\) 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

\[\text{cf-utt}(u, u')\]

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 \(\text{bond}(u, u')\) to indicate that \(u\) is bond to \(u'\).\(^{11}\)

\(^{10}\)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).

\(^{11}\)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.

After establishing accessibility through ACC-FROM-BOND, or possibly through other shallow coherence-inference methods, the resolution stage can begin and PARS2pro can be applied to identify the resource situation.

We can finally produce the revised version of PARS2pro 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 PARS2pro 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 $U_{NP}^1$ 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 $\text{do}$ of $U_{NP}^1$.

Let us now return to the data from Arnold and consider first (29a) uttered in the discourse situation following (28b), in which the two CFs are of different gender. As only mces $up1.1$ (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 (30). (The same happens when “she” is uttered, with both contexts in (28).)

(30) $[\text{bond}(up1.1, up2.1)]$

This hypothesis is immediately committed to, resulting in $\text{ACC-FROM-BOND}$ being triggered. This in turn leads to the following update:

(31) $[\text{accessible}(ce1, ce2)]$

which in turn triggers $\text{PARS2}_{pro}$, resulting in the interpretation in (32)

(32) $u_{NP}:\lambda P[y|y \text{ is } \iota x K1; ||x \text{ is } z||; P'(y)]$

$u_{Det}:\lambda P$

$sie: \lambda P[y|y \text{ is } \iota x K; P(x)]; P'(y)]$

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 (28a), in which both $up1.1$ and $up1.2$ match $up2.1$ in gender. As a result, $\text{PRO-MATCH}$ can be activated in two different ways, producing the two distinct hypotheses in (33)

(33) a. $[\text{bond}(up1.1, up2.1)]$

b. $[\text{bond}(up1.2, up2.1)]$

Each of these hypotheses in turn activates $\text{ACC-FROM-BOND}$—the updates resulting from this default are however identical (and identical with (31)). 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 $\text{PRO-MATCH}$ is at play in the case of first-mention entities. When this default, shown below and that we call $\text{PRO-MATCH-FM}$, is triggered, it overrides the weaker $\text{PRO-MATCH}$; otherwise a conflict between weaker defaults is obtained, which typically results in a toss-up between the alternatives.

\[
\begin{align*}
\text{cat}(U_{pro}) & \text{ is PRO,} \\
\text{cf-utt}(U_{n+1}, U_{pro}), & \\
\text{cf-utt}(U_{n}, U_{np}), & \\
\text{first-mention}(U_{n}, U_{np}), & \quad : \text{bond}(U_{np}, U_{pro}) \\
CU & \text{ is } U_{n+1}, \\
\text{agr-match}(U_{np}, U_{pro}), & \\
\text{bond}(U_{np}, U_{pro}), & \quad \frac{\text{PRO-MATCH-FM}}{}
\end{align*}
\]
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 (15). Upon hearing red, the discourse situation is updated with the expectation of encountering an N', as in (34). After parsing, this interpretation is adjoined to the syntactic interpretation of the, resulting in the updated interpretation in (36).

\[
\text{(34)} \quad [u_{\text{red}} : \text{utter}(A, "red")]
\]

\[
\begin{align*}
\text{red: } & \lambda P \lambda x[\text{red}(x)]; P(x) \\
u_{\text{red}} : & \text{Adj} \\
u_1^1 : & N' \\
u_2^1 : & N'
\end{align*}
\]

\[
\text{(35)} \quad [u_{N'}^1, u_{N'}^2]
\]

\[
\begin{align*}
\text{red: } & \lambda P \lambda x[\text{red}(x)]; P(x) \\
u_{\text{red}} : & \text{Adj} \\
u_1^1 : & N' \\
u_2^1 : & N'
\end{align*}
\]

\[
\text{(36)} \quad [\ ]
\]

The evidence from Eberhard et al. suggests that, at least in interpretive contexts like the visual world scenarios, updates like (36) are sufficient to trigger reference resolution. This translates into an hypothesis that a version of \text{PARS1} exists activated by the observation of an utterance of an adjective...
with semantic interpretation $\lambda P \lambda x [P(x); P(x)]$. This version of PARSi, that we call PARSi$_{Adj}$, is shown below. This version again requires a visual situation $K_{mvu}$ to be in the mutual focus of attention, but unlike in 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.

```
PARSi$_{Adj}$

$[U_{NP} : NP [Det \ the: \ \lambda P \lambda P'(\{y|y \ is \ \iota x P(x)\}; P'(y))]$

$[U_{N'} : N' \ [U_{1_{Adj}} : Adj: \lambda P \lambda x [[P(x)]; P(x)]

U_{2_{Adj}} : N'] ]$

MSOA is $K_{mvu}$

$K_{mvu} \models P(Z)$

$\Rightarrow$

$[U_{NP} : NP [Det \ the: \ \lambda P \lambda P'(\{y|y \ is \ \iota x P(x)\}; P'(y))]$

$[U_{N'} : N' \ [U_{3_{Adj}} : Adj: \lambda P \lambda x K_{mvu}: [[P(x)]; [\{x \ is \ Z\}; P(x)]

U_{4_{Adj}} : Adj: \lambda P \lambda x [[P(x)]; P(x)]] U_{2_{Adj}} : 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 PARSi$_{Adj}$ 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 PARSi—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.
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 MCIs, 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. 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.

12 Another being Combinatorial Categorial Grammar (Steedman 2001).
13 A more detailed comparison between PTT and Dynamic Syntax can be found in (Poesio and Rieser 2010).

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 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 rather 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.

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 that the present theory of belief revision should be extended to provide an account of reanalysis (Fodor and Ferreira 1998) and repairs (Ferreira et al. 2004).
Acknowledgments

We wish to thank the anonymous reviewers of this paper for many insightful comments.

References


Crain, S. and Steedman, M. (1985). On not being led up the garden path: the use of context by the psychological syntax processor. In D. R. Dowty,


Figure 4. The top panel shows samples for one-referent (pencil) and two-referent (apple on napkin) conditions. The bottom panel shows the proportion of looks to the competitor goal (the towel) for instructions with locally ambiguous and unambiguous prepositional phrases in one-referent and two-referent contexts.

Figure 1: The Visual World Paradigm: Ambiguous and unambiguous visual world situations
Figure 2: The materials used by Arnold et al. to study pronouns with the Visual World Paradigm