Utterance Processing and Semantic Underspecification

Massimo Poesio
University of Edinburgh
Human Communication Research Centre
E-mail: poesio@cogsci.ed.ac.uk

July 27, 1999

Abstract

We propose a theory of utterance processing meant to clarify the respective roles of incrementality and underspecification in semantic interpretation. After reviewing the available psychological evidence, we introduce (i) a theory of the semantic interpretations constructed by the language processor while processing utterances in an incremental fashion, including the cases in which these interpretations are semantically underspecified; and (ii) a formal model of the disambiguation process. We use our theory to account for the psychological findings concerning two areas of semantic interpretation: lexical disambiguation and pronoun resolution, emphasizing similarities and differences between the two processes. We also make a few preliminary hypotheses about scope disambiguation. The novel aspects of our proposal with respect to previous work on semantic underspecification include our emphasis on psychological evidence, our use of a standard logical formalism to characterize semantically underspecified interpretations rather than logics specially developed for this purpose, and the role in the theory of logics for defeasible reasoning, which are used to give a formal characterization of certain types of interpretations observed in psychological experiments.

1 Motivations

We believe with Frazier and Rayner (1990) that one can't make fully explicit hypotheses about semantic interpretation in the absence of "... a theory of the semantic decisions encountered by the human processor during language interpretation" (Frazier and Rayner, 1990, pg. 182). The results presented by Frazier and Rayner, and in other recent papers, suggest that so-called SEMANTICALLY UNDERSPECIFIED INTERPRETATIONS are a crucial ingredient of a psychologically motivated theory of semantic interpretation; but as of yet there exists no formally specified account of these results and how they can be reconciled with the finding from a number of experiments that certain aspects of semantic interpretation take place immediately (Swinney, 1979; Tanenhaus *et al.*, 1995). There seems to be a gap in this area between psychologically oriented research and research in computational semantics. Most psychologically motivated models of utterance processing concentrate on syntax (Sturt and Crocker, 1996; Jurafsky, 1996), or on the role of incrementality in disambiguation (Shieber and Johnson, 1993; Chater *et al.*, 1995). And although underspecified interpretations are increasingly viewed as a central ingredient both of theories of parsing motivated by psychological evidence (Gorrell, 1995; Frazier and Clifton, 1997; Sturt and Crocker, 1996) and of theories of semantic disambiguation (see, e.g., the papers in (van Deemter and Peters, 1996)), in computational semantics underspecification is seen as a computationally advantageous way of dealing with ambiguity and/or an interesting object of study from a logical perspective (van Deemter, 1991; van Eijck and Jaspars, 1996); motivating or evaluating these theories on the basis of psychological evidence has not been a priority.

Our aim in this paper is to fill this gap by providing a very general theory of semantic processing that reconciles the evidence about incrementality with the evidence that syntactically and semantically underspecified interpretations play a central role in language processing, without however abandoning the theories concerning the meaning of lexical items already developed in formal semantics. The reader won't find many innovations over previous work by the author and others as far as the general shape of theory is concerned, or the basic technical tools used. The theory of underspecification developed here builds upon ideas discussed in (Poesio, 1996b,c) and is related to proposals of other researchers, especially Muskens (1995a); the general hypotheses about disambiguation are derived from (Poesio, 1994, 1996a) and are similar to those made in (Hobbs *et al.*, 1993; Alshawi, 1992; Lascarides and Asher, 1993). (We discuss these and other proposals in Section §9.) The main contribution of this paper is our analysis of the psychological evidence about the respective roles of incrementality and underspecification in semantic disambiguation, and the way our technical tools are put to work to account for these data.

The main steps in building such a theory are (i) explaining how semantic interpretation can proceed incrementally, yet still in a compositional fashion ; (ii) explaining what it means for an interpretation to be (semantically) underspecified; and (iii) elucidating the respective roles of incrementality and underspecification in semantic interpretation–i.e., suggesting hypotheses, consistent with the known psychological evidence, as to when the interpretation of an utterance is underspecified, and when instead it is completely resolved. We will present our solutions to each of these problems in turn.

As for the applications of the theory, we concentrated on those aspects of semantic interpretation for which the psychological evidence is most clear: lexical disambiguation–formal accounts of which already exist, but that do not (to our knowledge) consider the results of Frazier and Rayner (1990); and the interpretation of pronouns, an area of semantic interpretation where a formal account of the psychological evidence we will discuss (Corbett and Chang, 1983; Garrod and Sanford, 1985; Garrod *et al.*, 1993; Poesio and Stevenson, To appear) is also missing. We also examine the (not very abundant yet) evidence that is available about scope assignment, and present a few preliminary proposals concerning this interpretation process. The model we propose is consistent with existing theories about the role of syntactic underspecification in parsing (Gorrell, 1995; Sturt and Crocker, 1996), and can be viewed as a extension of such proposals to cover semantic underspecification.

The content of the paper is as follows. We review the available evidence about incrementality and underspecification in Section §2. In Section §3 we propose our formal characterization of the semantic interpretations produced by the language processor,[1] and use this characterization to explain how semantic processing can proceed incrementally, and what it means for an interpretation to be (semantically) underspecified. In Section §4 we introduce a formal characterization of the process by which interpretations are produced, using lexical disambiguation as an example. In Section §5 we discuss anaphoric processing. In Section §6 we present a preliminary hypothesis on scope disambiguation, whereas in Section §8 we discuss

[1]We follow the terminology of Frazier and Rayner and use the term 'language processor' to refer to the language module of the entity that is processing an utterance.

some evidence that interpretations may remain underspecified even after a clause is completely processed. In the final sections we review related research, summarize our proposal and discuss the currently open issues in semantic underspecification.

2 Linguistic and Psycholinguistic Background

In this section we first summarize some assumptions of current work in formal semantics, and then review evidence about incrementality and underspecification in the psychological literature.

2.1 Theories of Semantic Competence

Formal semantics has developed detailed and remarkably successful analyses of a large variety of semantic phenomena (von Stechow and Wunderlich, 1991; Kamp and Reyle, 1993; Carpenter, 1998). These analyses are the natural starting point for a rigorous theory of semantic interpretation. Unfortunately, existing theories of semantic competence, just like theories of syntactic competence, have characteristics that make it difficult to extract theories of semantic interpretation from them. One of our goals will be to show that these problems can be corrected without abandoning the analyses proposed by semanticists.

Formal analyses of the meaning of a sentence do two things: they identify its possible interpretations, and specify how these interpretations can be derived in a compositional fashion by means of syntax-driven derivations which work bottom-up from the meaning assigned to that sentence's lexical constituents. We will illustrate this by discussing how a simple sentence, *Kermit croaked*, is analyzed. This sentence, under the assumption that *croaked* has two interpretations ('making a frog-like sound' and 'dying'), has two readings. (Here, and in the rest of the paper, we assume without loss of generality that these are the only two readings of the sentence.)

We use to characterize semantic interpretations Compositional DRT (Muskens, 1995b, 1994), which, of all formalisms currently used in semantics that do provide the necessary tools for an analysis of tense and anaphora, is the one that comes closest to a 'lingua franca'. Compositional DRT extends the type-theoretical logics traditionally used in formal semantics (Montague, 1973; Dowty *et al.*, 1981) by incorporating technical devices introduced in Discourse Representation Theory (DRT) (Kamp and Reyle, 1993) for specifying the truth conditions of sentences containing anaphoric expressions, as well as DRT's treatment of tense in terms of events and times; both of these extensions will be needed in the rest of the paper. A summary of the most important characteristics of Compositional DRT is in Appendix A.

In Compositional DRT, NPs are viewed as denoting sets of properties, following (Montague, 1973); and the semantic contribution of the n-th occurrence of the string *Kermit*[2] is to introduce a new discourse entity x_n denoting the object 'Kermit', to which the properties expressed by the rest of the sentence are to be applied; this is formally expressed by having the utterance of *Kermit* introduce a new Discourse Representation Structure (DRS), denoted by $[x_n|...]$, which is to be concatenated to the DRSs obtained by applying property P to x_n:

$$Kermit_n \rightsquigarrow \lambda P[x_n|x_n \textbf{ is } k]; P(x_n)$$

where x_n is a new discourse entity (which, in compositional DRT, are treated as constants of a special type, π_e), and k is a constant of type e denoting Kermit. (In what follows, the symbols x, x_i, y, and y_i will be reserved for discourse entities of type π_e.) Assuming a notion of lexical entry similar to those proposed in sign-based theories such as HPSG (Pollard and Sag, 1987, 1994) or Construction Grammar (Fillmore, 1988), we obtain for (the n-th utterance of) the string *Kermit* the lexical entry shown in (1), in which we have used a notation inspired by the one in (Pollard and Sag, 1987). This lexical entry associates to the string three types of information: the semantic translation in Compositional DRT, its syntactic properties, and phonological information:[3]

(1) $$Kermit_n \rightsquigarrow \begin{bmatrix} \text{phon} : \text{“kermit”} \\ \text{syn} : \begin{bmatrix} \text{cat} : \text{pn} \\ \text{agr} : \begin{bmatrix} \text{pers} : \text{sg} \\ \text{num} : 3 \end{bmatrix} \end{bmatrix} \\ \text{sem} : \lambda \text{P}[x_n|x_n \textbf{ is } k]; \text{P}(x_n) \end{bmatrix}$$

[2]In DRT, one has to distinguish between distinct utterances of the same string, because in general they may introduce distinct discourse entities.

[3]Some recent theories of the lexicon assume much richer lexical entries, including, e.g., argument structure (MacDonald *et al.*, 1994) or so-called *qualia* structure (Pustejovsky, 1991; Copestake and Briscoe, 1995; Lascarides *et al.*, 1996a). Although the type of lexical entries we eventually adopt in the paper is more complex than the example in (1), they will remain pretty simple in this respect.

We will assume throughout the paper that the verb *croaked* has the two senses mentioned before: 'to utter the sound that frogs utter' and 'to die', which we will express as the properties $\mathbf{croak}_f$ and $\mathbf{croak}_d$. Furthermore, we will assume that the two senses correspond to distinct lexical entries: i.e., we will treat *croaked* as a case of homonymy, like *bank*, rather than as a case of polysemy, like *newspaper*. (We return on this distinction below.) Compositional DRT incorporates a Davidsonian theory of events (Davidson, 1967), in which events, states and other so-called EVENTUALITIES are individuals in the universe of discourse that can be referred to, and predicates describing events or states, such as **croaked** or **love**, have an additional argument for the 'eventuality'. Following the convention adopted in DRT and Situation Theory, we will write e:$\mathbf{p}(\overline{x})$ to indicate the statement that an event e of type p, with remaining arguments the elements of vector $\overline{x}$, has taken place. A past tensed telic verb is then assigned a translation specifying that an event of that type occurred at a time before speech time S, as follows:[4]

$$\textit{croaked}_n \leadsto \lambda x[e_n | e_n{:}\mathbf{croak}_f(x), e_n < S]$$

The two lexical entries for *croaked*, then, are as in (2) and (3):

(2) $$\textit{croaked}_m \leadsto \begin{bmatrix} \mathsf{phon} : \text{“croaked”} \\ \mathsf{syn} : \begin{bmatrix} \mathsf{cat} : \mathsf{iv} \end{bmatrix} \\ \mathsf{sem} : \lambda x[e_m | e_m{:}\mathbf{croak}_f(x), e_m < S] \end{bmatrix}$$

(3) $$\textit{croaked}_m \leadsto \begin{bmatrix} \mathsf{phon} : \text{“croaked”} \\ \mathsf{syn} : \begin{bmatrix} \mathsf{cat} : \mathsf{iv} \end{bmatrix} \\ \mathsf{sem} : \lambda x[e_m | e_m{:}\mathbf{croak}_d(x), e_m < S] \end{bmatrix}$$

The interpretations of a sentence such as *Kermit croaked* are constructed in a compositional fashion, starting from the interpretation of its lexical constituents: each phrasal constituent gets an interpretation as the result of an application of a semantic operation to the interpretations of its constituents. For example, the derivation of the interpretation of the sentence according to which an event of Kermit performing a frog-like sound occurred at some time in the past is shown in (4).

[4] We have simplified matters considerably with respect to Muskens' treatment in (Muskens, 1995b); in particular, we have ignored the issue of the referentiality of times and reference time shifting.

(4)

$$\begin{bmatrix} \mathsf{syn}: \begin{bmatrix}\mathsf{cat}: \mathsf{s}\end{bmatrix} \\ \mathsf{sem}: [x_n, e_m | x_n \textbf{ is } k, e_m{:}\mathbf{croak}_f(x_n), e_m < S] \end{bmatrix}$$

$$\begin{bmatrix} \mathsf{syn}: \begin{bmatrix}\mathsf{cat}: \mathsf{np}\end{bmatrix} \\ \mathsf{sem}: \lambda \mathsf{P}[x_n | x_n \textbf{ is } k]; \mathsf{P}(x_n) \end{bmatrix} \qquad \begin{bmatrix} \mathsf{syn}: \begin{bmatrix}\mathsf{cat}: \mathsf{vp}\end{bmatrix} \\ \mathsf{sem}: \lambda x[e_m | e_m{:}\mathbf{croak}_f(x), e_m < S] \end{bmatrix}$$

$$\begin{bmatrix} \mathsf{phon}: \text{“kermit''} \\ \mathsf{syn}: \begin{bmatrix}\mathsf{cat}: \mathsf{pn}\end{bmatrix} \\ \mathsf{sem}: \lambda \mathsf{P}[x_n | x_n \textbf{ is } k]; \mathsf{P}(x_n) \end{bmatrix} \qquad \begin{bmatrix} \mathsf{phon}: \text{“croaked''} \\ \mathsf{syn}: \begin{bmatrix}\mathsf{cat}: \mathsf{iv}\end{bmatrix} \\ \mathsf{sem}: \lambda x[e_m | e_m{:}\mathbf{croak}_f(x), e_m < S] \end{bmatrix}$$

The other interpretation of the sentence,

(5) $[x_n, e_m | x_n \textbf{ is } k, e_m{:}\mathbf{croak}_e(x_n), e_m < S]$

is constructed in a similar way, but using the other lexical entry for *croaked*. More in general, each time a new reading is discovered, that reading can be characterized either by assuming a new lexical entry, or by hypothesizing a different syntactic analysis, or by associating additional semantic operations to the nodes of the parse tree. Of course, sentences like *Kermit croaked* could be analyzed equally well using simpler formalisms: e.g., by using first order logic as the meaning specification language and using unification to specify how meaning is composed (Pereira and Shieber, 1987). The motivations for the additional complexity of formalisms such as Compositional DRT and other formalisms derived from Montague's work can only be understood by looking at more complex examples. The problems with using first-order logic instead of the lambda-calculus for meaning composition are discussed in Moore (1989); as for the use of DRSs, the problem addressed by the theory is how contexts can be constructed incrementally– i.e., which discourse entities are made available by a sentence, to serve as potential antecedents of subsequent anaphoric reference. This problem is solved in DRT by assuming that the conventional meaning of utterances involves two parts: the discourse entities introduced by that utterance, and their properties (this is what a DRS is). Compositional DRT adds a sequencing operator ';' so that it's possible to construct the DRT representation of a complex text sentence by sentence. For example, (one) interpretation for the text in (6a) would be obtained by first constructing the interpretation for *Kermit croaked* as in in (4), then concatenating this interpretation to the DRS constituting the interpretation of *He is a frog*. (In which *He* is translated as introducing a discourse entity u_1 which is equal to x_n.)

(6) a. *Kermit croaked. He is a frog.*

b. $[x_n, e_m |\ x_n$ **is** k, e_m:$\mathbf{croak}_f(x_n)$, $e_m < S$];
$[u_1 | u_1$ **is** x_n, $\mathbf{frog}(u_1)]$

We will conclude this brief overview by pointing out that this example raises another issue to which we will return later on–namely, that one of the subtasks of semantic interpretation is to identify the antecedents of anaphoric expressions, but theories of semantic competence do not have much to say here, since we cannot expect the lexicon to specify all alternative interpretations of pronouns like *she*, *them* and *it* in the second sentence of (6a).[5] Such interpretations have to be extracted from the context, which means that a theory of semantic interpretation must also say something about the interaction between grammatical information and other sources of information used in interpreting utterances.

2.2 Turning a Theory of Semantic Competence into a Theory of Semantic Interpretation: Incrementality

In its simplicity, the examples discussed in the previous section already highlight the conflict between the goals of theories of semantic competence and theories of semantic interpretation. The goal of semantic theory is to explain how *all* interpretations of a sentence may be obtained; the question of which interpretation is preferred in a given context, or why, does not arise. If we were to see the formulation of the relation between expressions of English and their interpretations presented in Section §2.1 as an hypothesis about the respective roles of grammar and context in the interpretation of an utterance, we would get a theory of disambiguation as a 'generate and test' procedure, in which the role of linguistic knowledge is to generate all the readings of an ambiguous sentence–characterized, e.g., by means of a derivation tree annotated with expressions of the disambiguated language–whereas the role of context is to filter those which are implausible and / or choose the preferred one.

It has long been known, however, that it's not very likely that humans use generate-and-test to choose the intended interpretation of an utterance

[5]Although this is precisely what Montague assumed, since he was concerned with characterizing the space of logically possible interpretations rather than with developing a psychologically plausible theories of the lexicon.

in a given context. The number of distinct disambiguated interpretations of any given sentence is so large—scopally ambiguous sentences such as (7), for example, might have $n!$ readings, where n is the number of scope-taking operators—that it's not clear how humans could generate, let alone choose from, all these interpretations.

(7) *In most democratic countries most politicians can fool most of the people on almost every issue most of the time.*[6]

The problem of turning theories of linguistic competence into theories of language processing confronts syntactic theorists as well; and already in the '70s it was hypothesized that what prevented combinatorial explosion in the case of parsing was that the number of alternative interpretations was limited by the fact that interpretation took place incrementally. Primary evidence for this hypothesis was the phenomenon of GARDEN PATHS (Bever, 1970): sentences such as those in (8), 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 *raced*, *floated* etc. is immediately resolved in favor of the matrix verb interpretation, thus forcing the reader to a reanalysis step later on. These data led to the development of so-called GARDEN-PATH THEORY (Frazier, 1979, 1987) and numerous incremental models of parsing (Abney, 1991; Shieber and Johnson, 1993; Milward, 1994).

(8) a. *The horse raced past the barn fell.*
b. *The boat floated down the river sank.*

Both intuition and psychological evidence suggest that semantic interpretation is incremental, as well. First of all, semantic garden paths can occur, and are often used for humorous purposes. (9a) is an example of garden path relying on the lexical ambiguity of *strikes*; (9b) an example of garden path relying on the scopal ambiguity of *every 11 seconds a man is mugged here in New York City*; and the joke in Figure 1 relies on the fact that the pronoun *them* has more than one possible antecedent in that context, one of which is most preferred:

(9) a. *The first thing that strikes a stranger in New York is a big car.* (Raskin, 1985)
b. *Statistics show that every 11 seconds a man is mugged here in New York City. We are here today to interview him.*[7]

[6]This (mis)quotation from Lincoln is reported in Hobbs (1983).
[7]Reported by Barwise, as quoted in Milward and Cooper (1994).

Figure 1: Referential disambiguation and garden paths

Examination of the transcripts of spoken conversations also suggests that interpretation clearly begins before utterances have been assembled together in complete sentences. For instance, in the following fragment (from the TRAINS corpus of dialogues collected at the University of Rochester (Gross *et al.*, 1993)) it's clear that the repair in utterance 10.1 is 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).

(10)

```
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
```

These intuitions are supported by psychological evidence. For the case of lexical disambiguation, well-known cross-modal priming experiments (Swinney, 1979; Seidenberg *et al.*, 1982) indicated that although all senses of an ambiguous word were activated, all but the chosen one were immediately discarded. In these experiments, the subjects were presented with texts such as the one in (11); half of the time a disambiguating context was provided (the string *spiders, roaches and other*). Swinney found priming effects for both *ant* and *spy* at [1], even with a strongly disambiguated context; but only for *ant* at [2].

(11) *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.*

Eye-tracking experiments by Tanenhaus *et al.* (1995) showed that definite descriptions are also resolved incrementally, at least in the case of visually presented antecedents. Indeed, the extreme sensitivity of the measuring mechanism used allowed Tanenhaus and colleagues to ascertain that definite descriptions were resolved as soon as a disambiguating substring of the head noun was perceived–thus, in a visual context containing candies, candles and forks, the instruction *Pick up the candle* was processed more slowly than the instruction *Pick up the fork*, because the visual context contains an object with a prefix in common with 'candle', the candies. Tanenhaus et al also confirmed that this early processing of semantic information affected syntactic disambiguation, as already suggested by Crain and Steedman (1985). [8]

The conclusion that we can draw from these studies is that we need a theory of how semantic processing may occur in an incremental fashion. The problem is that the theories of semantic competence discussed in Section §2.1 do not assign to lexical items propositional translations, i.e., objects which can be true or false. Such theories cannot explain, therefore, how the inferences that result in an interpretation being chosen incrementally (as in (10)) can take place, because the interpretation obtained as a result of perceiving a new word cannot be characterized simply as the conjunction of the previous interpretation and the lexical translation of the new word; most of the times, this lexical translation will not be an object that can be conjoined. This problem has been called by Chater *et al.* (1995) the PARADOX OF INCREMENTAL INTERPRETATION, and is of special concern when

[8] In the case of scope disambiguation the evidence is less clear. We will return to this topic in Section §6.

trying to account for how semantic interpretation (e.g., anaphora resolution) takes place. In order to develop a theory of semantic interpretation consistent with these theories of semantic competence, it is necessary to address this problem.

2.3 Parallelism

One of the most debated questions about language processing is whether during interpretation only one hypothesis at a time is considered (SERIAL interpretation) or whether multiple interpretations are considered in parallel. This debate is still ongoing, but more and more evidence suggests that certain disambiguation processes, at least, involve parallel generation of more than one hypothesis.

In the case of lexical access, evidence supporting the claim that all meanings of a word are accessed in parallel was presented in (Swinney, 1979; Tanenhaus *et al.*, 1979; Seidenberg *et al.*, 1982; Marslen-Wilson, 1984). Since then, new results (Rayner and Morris, 1991; Tabossi, 1988) have complicated the picture somewhat; currently, the most plausible account of the data seems to be that all lexical entries associated with a word-string are activated in parallel, but the degree of activation is sensitive to the relative frequencies of the meanings and the context in which the ambiguous words occurs (for a review, see (Simpson, 1994); also (Hirst, 1987), chapter 4).

Evidence that multiple hypotheses are entertained in the case of pronoun interpretation, as well, is discussed in (Corbett and Chang, 1983; Stevenson and Vitkovitch, 1986; Gernsbacher, 1989). In Corbett and Chang's experiments, for example, subjects were presented with sentences such as:

(12) a. *Karen poured a drink for Emily and then Karen put the bottle down*
b. *Karen poured a drink for Emily and then she put the bottle down*

and then with a probe word referring either to the antecedent of the subject of the second clause, or to the non-antecedent. Corbett and Chang found that the non-antecedent (*Emily*) was much more activated when a pronoun has been used to refer to *Karen* (as in (12b)) rather than a proper name (as in (12a)).

In the case of parsing, the evidence is less clear: for many years, garden paths were taken as evidence that parsing is a serial process (Frazier, 1979; Marcus, 1980). More recently, however, several researchers have proposed accounts of garden paths in terms of parallel models in which alternative hypotheses are pruned (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). Parallel models of parsing include (MacDonald *et al.*, 1994; Stevenson, 1994; Jurafsky, 1996).

2.4 Underspecification

In the past years, the picture presented in Section §2.2 has begun to change as evidence emerged that although language interpretation is largely incremental, some decisions may be deferred. In some cases, partial interpretations are initially constructed, and perhaps refined later; in other cases, more than one hypothesis is kept around until the end of the sentence. We review this evidence in this section.

Underspecification in Parsing

In the case of parsing, what needs explaining is why certain types of garden paths are 'harder' than others. It was noted that whereas sentences such as *While John walked the dog barked* seem to cause genuine difficulties for readers, sentences which should equally involve a reanalysis process, such as *John knows Mary is smart* (for which one would expect an initial preference for *Mary* being interpreted as object of *knows*, given the relative frequency of the two constructions), in fact do not seem hard to parse (Frazier and Rayner, 1987; Gorrell, 1995; Sturt and Crocker, 1996; Frazier and Clifton, 1997). This difference between 'hard' and 'easy' cases of reanalysis was explained by Gorrell (1995); Weinberg (1993); Sturt and Crocker (1996) by suggesting that the parser initially builds 'partial' or 'underspecified' interpretations; these interpretations make some types of reanalysis easy, and other hard.

Most theories of parsing involving underspecified interpretations build on DESCRIPTION THEORY, proposed by Marcus *et al.* (1983) and axiomatized, e.g., by Backofen *et al.* (1995). According to this theory, syntactic interpretations can be expressed in terms of *indirect* as well as direct constituency relations: i.e., it is possible to state that a certain phrase XP is either a direct constituent of another phrase YP, or that there is a chain of intermediate constituents such as XP is a constituent of the first of them and the last one is an immediate constituent of YP. The parsing theorists mentioned above propose that in some cases, the initial syntactic interpretation of a sentence only involves such indirect constituency relations: for example, the initial interpretation of *I saw a man with a telescope* could be one in which the attachment positions of *I* and *saw*, and the position of *a telescope* within the PP *with a telescope*, have been completely determined; but for the other sentence constituents, only partial information is available–namely, that the NP *a man* and the PP *with a telescope* are subordinate to the VP, but nothing more. This interpretation is illustrated in Fig. 2.

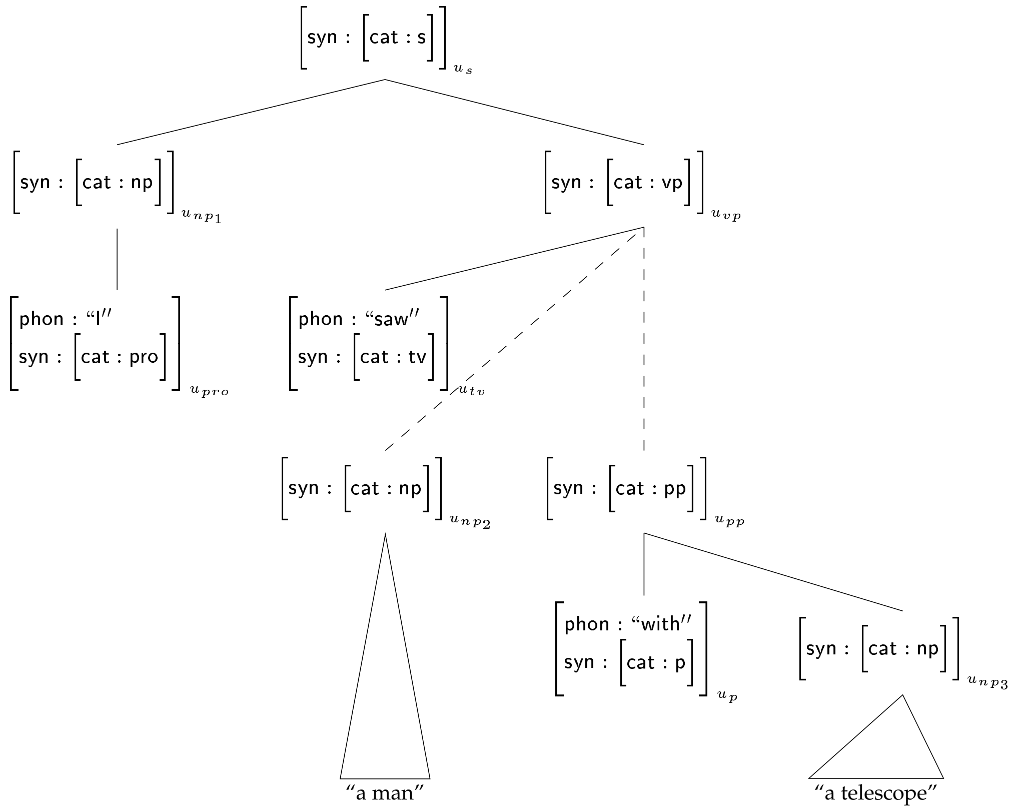


Figure 2: A syntactically underspecified interpretation of *I saw a man with a telescope*.

Parsing theories that assume such interpretations can distinguish between two types of reanalysis: 'easy' cases, that only involve adding information about constituency to the present interpretation; and 'hard' ones, that involve revising existing information. Thus, for example, reanalyzing *John knows Mary is smart* as a case of sentential complement only involves adding an S node dominating the NP *Mary*; under the assumption that the NP is originally interpreted as being indirectly dominated by the VP with head *knows*, this addition which does not involve any structural modification to the previous interpretation. By contrast, reanalyzing *While John walked the dog barked* or *The horse raced past the barn fell* involves revising some decisions about constituency that had been already made; this is what makes these cases 'hard'.

Current theories of the role of syntactic underspecification in parsing differ in the details concerning the conditions under which reanalysis is 'easy'. We will follow here Sturt and Crocker (1996), who use partial tree representations based on (Lexicalised) Tree Adjoining Grammars (Schabes, 1988) and assume that a set of four operations can always be performed on partial syntactic interpretations without involving additional cost: Left and Right Attachment, TAG Adjunction, and Tree Lowering.

Underspecification in Semantic Interpretation

Recent experiments support the view that semantically underspecified interpretations play in semantic interpretation a role similar to that of syntactically underspecified interpretations in parsing: i.e., that in certain cases the decision concerning the semantic interpretation of a constituent may be postponed.

Lexical Disambiguation

Frazier and Rayner (1990) contrasted what they call the IMMEDIATE COMPLETE INTERPRETATION HYPOTHESIS, according to which the language processor interprets each phrase fully as soon as it is encountered, with what they call the IMMEDIATE PARTIAL INTERPRETATION HYPOTHESIS:

> **Immediate Partial Interpretation Hypothesis (IPIH)**: The processor may delay semantic commitments, if this does not result in either (a) a failure to assign any semantic value whatsoever to a word or major phrase, or to its relation to other major phrases or (b) the need to maintain multiple incompatible values for a word, phrase or relation.

In order to evaluate these two hypotheses, Frazier and Rayner used materials including two types of words: words 'with two meanings', i.e., homonyms such as *pitcher* or *croaked*;[9] and words 'with two senses', i.e., polysemous words such as *newspaper* which have been argued to have a single lexical entry associating to them an interpretation which is subsumed by

[9] Homonyms are also called HOMOGRAPHS.

all of their senses (Cruse, 1995; Pinkal, 1995). Frazier and Rayner hypothesized that in the case of words with two lexical entries, subjects would have to make an early commitment, whereas with words with a single entry but two senses, the more general interpretation associated with the lexical entry could be adopted initially, and would only be refined at the end of the sentence. This, they predicted, would result in garden paths in appropriate contexts with 'two-meaning' words, but not with 'two-senses' ones.

Frazier and Rayner's materials included three sets of sentences: Sets 1 (examples of which are in (13)) and 2 (in (14)) contained ambiguous words with two meanings; Set 3 (in (15)) contained ambiguous words with two senses. In all cases, the clause that contained the target word was neutral with respect to its two interpretations; each set contained both sentences in which the disambiguating clause occurred before the clause with the target word (a-b), and sentences in which the disambiguating clause occurred after (c-d). Frazier and Rayner established before the experiments what was the preferred interpretation of each target word (i.e., how these words were interpreted when no disambiguating context was given); this, they assumed, would be the initial interpretation assigned to these words by their subjects when processing sentences in which the disambiguating clause followed. They predicted a garden path effect when a subsequent disambiguating clause suggested a non-preferred interpretation.

(13) a. *Being so elegantly designed, the pitcher pleased Mary.*
b. *Throwing so many curve balls, the pitcher pleased Mary.*
c. *Of course the pitcher pleased Mary, being so elegantly designed.*
d. *Of course the pitcher pleased Mary, throwing so many curve balls.*

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

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

The results confirmed Frazier and Rayner's hypothesis. Garden-paths effects were observed with sentences in Sets 1 and 2 in which the disambiguating clause followed the clause containing the target word, and the disambiguating clause suggested the non-preferred interpretation; but no such effects were observed with sentences in Set 3, containing 'two-sense' words.

Anaphora Resolution

Garrod and Sanford (1994) discuss experiments addressing similar issues for the case of anaphoric interpretation. In Garrod and Sanford (1985), detection of spelling errors was used as a diagnostic indicating whether anaphoric expressions had been interpreted, using materials such as:

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

a. *Within seconds, Elizabeth* **senk** *(sank) into the pool.*
b. *Within seconds, Elizabeth* **jimped** *(jumped) into the pool.*
c. *Within seconds, the lifeguard* **senk** *(sank) into the pool.*
d. *Within seconds, the lifeguard* **jimped** *(jumped) into the pool.*
e. *Within seconds, he/she* **senk** *(sank) into the pool.*
f. *Within seconds, he/she* **jimped** *(jumped) into the pool.*

The continuation sentences (a-f) reported actions that were consistent with what was known about only one of the characters in the story. Garrod and Sanford measured the time it took subjects to detect the spelling error in the relevant verb, expecting that it would take longer for them to detect the error with verbs describing actions not consistent with the rest of story. Indeed, this was the pattern observed with definite descriptions and proper names; but when *Elizabeth* and *the lifeguard* were replaced with the pronouns *he* and *she*, differences in error detection time were only observed when the pronoun referred to the main character of the story (*Elizabeth* in this example): when the pronoun referred to the lifeguard, no difference was found. The study was repeated using an eye-tracking procedure by Garrod *et al.* (1993), who observed a similar effect: again there was evidence for early detection of inconsistency, but only when the pronoun was unambiguous with respect to gender and maintained reference to the topic of the passage. Garrod and Sanford explain these results by proposing that whereas definite descriptions and proper names are immediately interpreted, pronouns are only immediately interpreted when they are unambiguous and refer to the topic of the passage.

These results suggest that in addition to the two disambiguation patterns identified by Frazier and Rayner–generation of all interpretations followed by immediate choice of an unambiguous conventional meaning for homonyms, or immediate choice of a partially specified conventional meaning for polysemous words–there is another option, at least in the case of pronouns: that the hypotheses generated by the processor may not be immediately resolved if there isn't sufficient disambiguating evidence.[10] In fact, it's not clear whether the subjects in the experiments reported by Garrod and Sanford attempted to construct a complete conventional meaning for some of the sentence at all. In other words, these results suggest that semantic interpretation may not always be carried to the very end; as a result, even the weakening of the Immediate Complete Interpretation Hypothesis suggested by Frazier and Rayner (the IPIH) may still be too strong a characterization of how much semantic processing takes place. (More evidence concerning this will be discussed in Section §8.)

Scope Assignment

The psychological evidence is less clear for other types of semantic processing, and in particular, for the interpretation process that originated most work on semantic underspecification, scope assignment. Intuitively, it's clear that scope assignment can't be completely incremental, at least if we see the problem as that of determining what the arguments of an operator (e.g., a quantifier or a modal expression) are: this is because although the restriction of the quantifier *Every kid* in (17) can be identified as soon as the NP is processed, its nuclear scope can only be determined after the whole VP *climbed a tree* has been interpreted.

(17) *Every kid climbed a tree*

Indeed, recent work on the semantics of focus indicates that sometimes even the restriction of adverbial quantifiers such as *always* can't be determined until later in the sentence. (18a), for example, asserts that whenever officers escorted someone in Saint Petersburg, it was ballerinas that they escorted; but (19a) asserts that whenever someone escorted ballerinas in Saint Petersburg it was officers who escorted them (Rooth, 1985). In this second case, whose logical structure is paraphrased in (19b), the restriction of the adverbial quantifier **always** can only be determined after *ballerinas* is encountered.

(18) a. *In Saint Petersburg, officers always escort [balleRInas]$_F$.*

[10] Frazier and Rayner's materials did not include cases in which no disambiguating clause was present.

b. **always**(*officer-escort-someone-in-S.Petersburg*)
(*person-escorted-is-ballerina*)

(19) a. *In Saint Petersburg, [OFficers]$_F$ always escort ballerinas.*
b. **always**(*someone-escort-ballerina-in-S.Petersburg*)
(*person-escorting-is-officer*)

Finally, Ioup (1975) showed, *contra* Lakoff (1971), that the relative scope of the operators in a sentence is not always determined by left-to-right order: e.g., whereas the preferred reading of (20a) is the one in which *all children* takes narrow scope with respect to *a picture*, as predicted by left-to-right order, in the case of (20b) the preferred reading is the one in which *each child* takes wide scope:

(20) a. *I showed a picture to all children.*
b. *I showed a picture to each child.*

These data led many computational semanticists to hypothesize that perhaps scope assignment is unlike other types of semantic processing in that it only takes place on demand (Poesio, 1991; Reyle, 1993). Other researchers proposed that scopally ambiguous sentences were among the cases of semantic interpretation to which Frazier and Rayner's Immediate Partial Interpretation Hypothesis applies, in that they initially got an underspecified interpretation that subsumed all readings (Kempson and Cormack, 1981; Kálmán, 1990). However, the psychological evidence runs counter both of these suggestions. The claim that scope disambiguation does not take place at all rests uncomfortably with the results reported in Kurtzman and MacDonald (1993), who showed that at least in the case of sentences such as (17), containing one existential quantifier and one universal quantifier (in various orders), subjects did produce a disambiguated interpretation, and this interpretation wasn't always what we would expect if disambiguation had taken place on demand. The results of Kurtzman and MacDonald are also problematic for the idea that scopally ambiguous sentences initially receive an interpretation which subsumes all others, since they found that the preferred interpretation of sentences such as *A kid climbed every tree* was the strongest one (e.g., the one in which *A kid* takes widest scope). Another problem for the claim that scopally ambiguous sentences initially receive an interpretation that subsumes all others is that although such 'most general interpretations' exist in the case of sentences containing only universal and existential quantifiers, such as (17) or the sentences in (20), they are not available in general–e.g., for the well-known example *Few students speak many foreign languages*.

Unfortunately, even the experiments of Kurtzman and MacDonald do not give us enough evidence to develop a theory of the respective role of incrementality and underspecification in scope assignment, since they don't tell us anything about the on-line timing of scopal interpretation. Furthermore, we do not know what happens with sentences containing more than two quantifiers, such as (7). For this reason, we will postpone discussing scope disambiguation until Section §6.

2.5 Ambiguity doesn't always go away

> *The warning message we sent the Russians was a calculated ambiguity that would be clearly understood. [Alexander Haig]*

From what we have seen so far it would seem that people always end up assigning a single interpretation to utterances. As noted in (Poesio, 1996a), this is not the case, however; in fact, it has long been known in the field of Rhetoric that it is possible to construct sentences in such a way as to make them ambiguous:

> The old Rhetoric (e.g., Aristotle) treated ambiguity as a fault in language, and hoped to confine or eliminate it. The new Rhetoric sees it as an inevitable consequence of language and as an indispensable means of our most important utterances - especially in Poetry. (Richards, 1936)

The semantic garden paths examples in (9) are examples of ambiguity exploited for humorous purposes; the excerpt below, from Hopkins' "The Wreck of the Deutschland" (reported in (Su, 1994, page 22)) is an example of lexical ambiguity exploited for poetical purposes. The topic of the poem is the wreck of a ship, in which five nuns gave up their places in the life boats to save some other passengers. In so doing, the poet tells us, they drowned (were *sealed* by waters), but they also gained Heaven (they were *sealed* by God's mark of approval). The passage can only be properly appreciated if both readings of *sealed* are perceived simultaneously.

(21) *and these thy daughters*
And five-lived and leaved favour and pride,
Are sisterly sealed *in wild waters,*
To bathe ...

2.6 Summary

The evidence discussed in this section gives us a rather complex picture of semantic processing. Semantic interpretations are produced incrementally, but this doesn't always result in the immediate determination of a unique interpretation. Sometimes multiple interpretations of a word are generated; other times a single interpretation that subsumes all the readings is first obtained; and in some cases of anaphoric processing studied by Garrod and Sanford, no interpretation appears to be produced at all. When multiple interpretations are generated, sometimes all but one get pruned immediately, as in the cases of lexical disambiguation of 'two-meanings' words studied by Swinney and Frazier and Rayner; in the cases of ambiguous pronouns studied by Corbett and Chang, however, the two interpretations remain active until the end of the sentence, and get pruned then; finally, in examples such as (21), the interpretations appear to remain active even after the end of the sentence. The possible scenarios for semantic interpretation can then be summarized as follows:

- the 'multiple-meanings' case studied by Swinney, in which multiple interpretations are generated in parallel, but all but one get eliminated right away;
- the variants of this scenario observed by Corbett and Chang, in which the multiple interpretations remain active until the end of the sentence, or even after that;
- the 'multiple-senses' case studied by Frazier and Rayner, in which only one interpretation is generated immediately, and refined at the end of the sentence;
- the 'no interpretation' case reported by Garrod and Sanford, in which apparently no semantic interpretation at all is produced.

This discussion indicates the need for a theory of semantic underspecification that may support research in semantic interpretation in the way that Marcus *et al*'s d-theory has supported work on syntactic underspecification. Such theory should provide tools to characterize both the cases in which the initial interpretation is underspecified and the cases in which multiple hypotheses are initially generated. Providing such a theory is the main goal of this paper.

3 The Semantic Interpretations Produced by the Language Processor

3.1 Interpretations as Information About Utterances

Our approach to the problem of characterizing the (possibly, underspecified) interpretations produced by a language processor behaving as described in Section §2 is similar to that taken in theories of parsing based on d-theory, such as (Sturt and Crocker, 1996): we used a logic to characterize in a formal way the information that the processor has at any time during interpretation. This approach is not just dictated by the desire to have a clean theory; in the case of semantic interpretation, a logic-based characterization of interpretations is virtually unavoidable, given that pretty much all large-coverage theories of semantic competence are based on logical formalisms, as seen in Section §2.1.

Our characterization of the semantic interpretations produced by the language processor is motivated by the belief that it is only possible to account for incremental processing and make sense of the notion of semantic underspecification by making explicit and formalizing a number of often unstated assumptions about meaning and the interpretation of utterances in context on which formal semantics relies.[11] A simplified characterization of such assumptions goes as follows. Assumption number one is that the goal of utterance interpretation is to identify the conventional meaning (in Grice's sense (Grice, 1957)) of an utterance in the given context, and this conventional meaning can be characterized using theories such as those proposed in formal semantics; we will use Compositional DRT for this purpose. Assumption number two is that whereas *sentences* generally have more than one meaning, *utterances* have exactly one conventional meaning in a given context. (We will not be concerned with what Grice called 'speaker's meaning' or 'meaning-nn' (Grice, 1957).) For example, we assume that an utterance of *Kermit croaked* in a given context has exactly one conventional meaning, that this conventional meaning is one of the two propositions discussed in Section §2.1, and that the goal of semantic interpretation when processing that utterance is to identify the conventional meaning of that utterance in the discourse situation.[12]

[11] The particular assumptions on which we depend have been spelled out in most detail by Lyons (1995).

[12] We are of course aware that in doing this we ignore complexities such as the fact that recognizing the speaker's intentions usually involves more than just identifying one meaning licensed by the grammar—indeed, that identifying the conventional meaning may not

These first two assumptions can be made more precise in terms of a theory of utterances and their meaning reminiscent of that proposed in Situation Semantics (Barwise and Perry, 1983). As in Situation Semantics, we assume that utterances are events (the LOCUTIONARY ACTS of speech act theory (Austin, 1962)); but unlike most formalizations of Situation Semantics, we assume a Davidsonian theory of events (Davidson, 1967). Following the convention introduced in the previous sections, we will write e:$\mathbf{p}(\overline{x})$ to indicate the statement that an event e of type p with remaining arguments the elements of vector $\overline{x}$ has taken place. We characterize utterances by means of the three-place predicate **utter**, whose arguments denote the agent, the string uttered, and the uttering event itself. Thus, we characterize a situation in which individual a utters the sentence-string *Kermit croaked* as one in which an utterance event u occurred, such that (22) is the case:

(22) u:**utter**(a,*"Kermit croaked"*)

We implement the assumption that each utterance has exactly one conventional meaning by hypothesizing a total function from utterance eventualities to their conventional meaning, which we indicate as $\rightsquigarrow$. We use this function to describe the situation in which utterance event u in (22) has been assigned the conventional meaning in (4) (that an event e occurred of object k ('Kermit') having the property $\mathbf{croak}_f$) as follows:

(23) u:**utter**(a,*"Kermit croaked"*) $\wedge\ u \rightsquigarrow [x_n,e_m |\ x_n$ **is** $k,\ e_m$:$\mathbf{croak}_f(x_n),\ e_m < S\ ]$

We make two further assumptions. Assumption number three is that the conventional meaning of utterances is derived in a compositional fashion from the conventional meaning of their constituents, starting from the conventional meaning of the utterances of lexical items, just like in theories of semantic competence such as those discussed in Section §2.1. Finally, assumption number four is that assigning the conventional meaning to any given utterance is just a part of the process of constructing the context in which this utterance has to be interpreted, and this process of context construction works much as proposed by Lewis and Stalnaker (Stalnaker, 1973; Lewis, 1979).

The connection between these assumptions and our stated goal of characterizing the semantic interpretations produced by the language processor is as follows. We propose that these interpretations do not abstract away

always be a prerequisite for identifying the speaker's meaning. We are developing a more general theory of interpretation in which some of these complexities are taken into account– see, e.g., (Poesio and Traum, 1997).

from the fact that an utterance has taken place, to consist only of information about the conventional meaning of the utterance. Instead, we hypothesize that a central property of these interpretations is that they are 'about' the occurrence of utterances: i.e., that the hypotheses about syntax and semantics that the language processor constructs are hypotheses about certain properties of these utterances. Because of this, the language processor knows when some information about a given utterance is missing; in fact, this knowledge plays a central role in guiding its inference processes. In the rest of this section we are going to make this hypothesis clearer, and to show that once we accept this hypothesis it becomes very easy to explain (i) how semantic interpretation can proceed incrementally without abandoning the theories about the conventional meaning of words generally accepted in formal semantics and (ii) what it means for an interpretation to be underspecified; and all of this without need for a new logic. We will do this in Sections §3.2 and §3.4, respectively.

It may seem, however, that this hypothesis goes counter a number of standard assumptions in formal grammar, since grammatical competence and its role in language processing, are typically specified abstracting away from utterance occurrence. We don't believe this is the case. In order to specify a theory of the interpretations entertained by the language processor in terms of utterances it is necessary to state how grammatical information is related to utterances; but although our way of formulating this association (indeed, the very idea of putting things this way) may look a bit unusual, we trust that what we are going to propose will not seem particularly problematic; in fact, we will see in a moment that what we are proposing is fairly similar to what a number of current theories of grammatical competence do. We spell out the connection in more detail in the rest of this section.

We have just introduced almost all the formal ingredients that will be required by our theory of the interpretations entertained by the human processor at various stages of interpretation. In particular, only very little more will be needed to characterize most types of semantically underspecified interpretations, except for scope (to which we return below). This means that the semantic interpretations we propose can be characterized using the formal tools provided in Compositional DRT, which we already planned to use to characterize the conventional meanings of lexical items. Compositional DRT is especially suited in light of our assumption number four, which can also be seen as telling us something about the nature of the semantic interpretations constructed by the language processor–namely, that these interpretations are partial representations of a context, incrementally updated as the result of new input and of inference. Interpretations of this type are naturally modelled as DRSs, as we will see in a moment.[13] We will not therefore introduce a special logic to characterize interpretations; we will only make a few extensions to Compositional DRT, discussed in Section §3.5.

3.2 Incremental Semantic Interpretation

One of the reasons for proposing the theory of semantic interpretations in Section §3.1 is to explain how semantic processing can take place incrementally, but at the same time proceed in the compositional fashion assumed by the theories of semantic competence discussed in Section §2.1.

We have seen how accounting for incremental processing is a problem for traditional theories of semantic competence, in that these theories propose non-propositional translations for lexical items, which therefore could not support the inference processes involved in disambiguation. Some theories of interpretation, most notably (Hobbs *et al.*, 1993), solve the problem by assigning to lexical items translations that are radically different from those proposed by most semantic theories. We show in this section that this is not necessary. The problem disappears once we take semantic interpretations to be characterizations of the properties of utterances; for in this case, the information that gets added to an interpretation when we discover the conventional meaning of a word *is* a proposition–the proposition that the utterance of that word has that meaning. In this section we will give a very simple illustration of what we mean by this, i.e., how the perspective on interpretation we have just presented can be used to specify the interpretations obtained while processing the input word by word in the case of the sentence already discussed in Section §2.1, *Kermit croaked*. We will do this by showing how various modules of the language processor update semantic interpretations, without however giving a formal characterization of the rules that produce such updates; those will be discussed in Section §4.

We assume, on the basis of the evidence about parallelism discussed in Section §2, that at any given moment the language processor entertains a number of competing interpretations of the utterances perceived until that moment–i.e., of the 'conversational score' in Lewis and Stalnaker's

[13]Indeed, we also used Compositional DRT to give a more general account of the dynamics of the common ground and how it gets established (Poesio and Traum, 1997; Poesio and Muskens, 1997).

sense. Each of these interpretations is a complete representation of a context, including both information gained through perception and information gained through inference, and is therefore best characterized as a Discourse Representation Structure (DRS) K in the sense of Kamp and Heim, containing information as in (23).[14] As a new utterance is perceived, each of the DRSs specifying an interpretation currently being entertained is updated with the information that that utterance occurred, thus obtaining new interpretations. This update can be formally viewed as a concatenation of the DRS representing the new information to the existing interpretation.

For example, assume that (one of) the current interpretations of the set of utterances perceived before hearing the utterance *Kermit croaked* (and of the information derived by lexical access, parsing, etc.) is the DRS K. Then perceiving an utterance of *Kermit* results in a new interpretation, which can be characterized in formal terms as the DRS K' obtained by concatenating K with a DRS representing the perceptual information in (22) about the new utterance, as follows:

(24) $K' = K; [u_{pn} | u_{pn}$:**utter**(a,"*Kermit*")]

K' gets then extended as the result of the interpretive processes involved in language interpretation. We consider each of these in turn.

Lexical Access

We have seen in Section §2.1 simplified examples of the type of information stored in the lexicon according to sign-based theories. We can turn these theories into theories of how the lexicon contributes to utterance interpretation by viewing lexical access as the process by which semantic interpretations such as K' in (24) are enriched by hypothesizing further properties of utterances such as u_{pn}. Assuming that the lexical entry for *Kermit* is as in (1), lexical access can then be seen as resulting in the update in (25):[15]

(25) [u$_{pn}$ | u_{pn}:**utter**(a,"*Kermit*"), $u_{pn} \rightsquigarrow \lambda P[x_n | x_n$ **is** $k]; P(x_n)$, **syn**(u_{pn}) **is** [cat : pn]]

[14] Such an interpretation encodes the proposition that there are assignments satisfying that DRS, $\exists ij K(i)(j)$; in what follows, however, we will omit the existential quantification over assignments and just use a DRS to characterize the current interpretation.

[15] The relation between the feature-based notation in (1) and the utterance-based notation in kermit:lex can be made more precise by specifying a mapping τ from signs into DRSs as follows. Let PHON be the interpretation of the attribute `phon`, SYN be the interpretation of `syn`, SEM the interpretation of `sem`, and DAUGHTERS the interpretation of the feature `daughters` specifying the subconstituents of a sign. Then, the interpretation of a sign as a DRS is specified by the following definition:

((25) is based on the assumption that part of the information contributed by the lexicon towards the interpretation of a lexical utterance is its syntactic classification; this assumption is formalized by the function **syn**, that maps utterances into the kind of objects that are assumed as the interpretation of signs of type SYN in sign-based theories such as HPSG *circa* (Pollard and Sag, 1987; Carpenter, 1992).) Updating K' with (25) originates a new interpretation characterizable as the new DRS K'':[16]

(26) $K'' = K'; [|u_{pn} \rightsquigarrow \lambda P([x_n|x_n \textbf{ is } k]; P(x_n)), \textbf{syn}(u_{pn}) \textbf{ is } \mathsf{cat} : \mathsf{pn}]$

We give more examples of lexical access in Section §3.4 (discussing more specifically underspecified lexical entries), and discuss the nature of the rules that result in updates such as (26) in Section §4.

Parsing

The updated interpretations resulting from lexical access serve as the basis for a second interpretation process, parsing. As syntactic processing is not the focus of this paper, we only include in this paper a minimal discussion of how syntactic information is represented and of our assumptions about how it is updated, mainly derived from (Sturt and Crocker, 1996).

The role of syntactic knowledge can be characterized within the present framework by viewing parsing as the generation of hypotheses about how utterance events are assembled together in larger utterance events. So, in the example we have been discussing, we can characterize the interpretation of the sentence-string *Kermit croaked* in (4) by saying that the language processor has constructed an interpretation according to which the utterance event u_{pn} is the single element of the decomposition of a larger utterance event u_{np} which, together with a second utterance event u_{vp}, is the

---

$$\tau\left(\begin{bmatrix} \mathsf{phon} : \mathsf{X} \\ \mathsf{syn} : \mathsf{Y} \\ \mathsf{sem} : \mathsf{Z} \\ \mathsf{daughters} : \left\langle [\ldots]_{u_1}, \ldots [\ldots]_{u_n} \right\rangle \end{bmatrix}_u\right) = \begin{array}{l} [u|u{:}\textbf{utter}(a, PHON(u)), \textbf{syn}(u) = SYN(u), u \rightsquigarrow SEM(u)]; \\ \tau([\ldots]_{u_1}); \ldots, \tau([\ldots]_{u_n}); \\ [|u_1 \lhd u, \ldots, u_n \lhd u] \end{array}$$

The interpretation of feature-based theories in terms of first-order logic has been specified in greater detail, e.g., by Johnson (1988).

[16] According to the 'cohort' theory (Marslen-Wilson, 1984) and most other recent theories of lexical access, the alternative lexical interpretations of a word are actually accessible before the entire word has been perceived, which means that a listener never entertains an interpretation like (24), but only underspecified interpretations about phonemes, such as u_1:**utter**(*a*,*"cro"*). For simplicity, we assume in this paper that the lexicon is not accessed until a complete word-string has been encountered; moving to a more accurate characterization would simply involve assuming a finer grained incrementality–e.g., one in which inference processes start after every phoneme.

decomposition of the larger utterance event u, classified as being of type S.[17]

This means that parsing, as well, can be cast in terms of updates to the current interpretation. We use the notation $u_1 \lhd u_2$ to indicate that event u_1 is part of the decomposition of event u_2, and the notation $u_1 < u_2$ to indicate that event u_1 precedes event u_2; with this notation, hypothesizing that u_{pn} is a constituent of an NP u_{np} would result in the following update:

(27) $K''' = K''; [u_{np} | \mathbf{syn}(u_{np}) \text{ } \mathbf{is} \text{ } [\mathsf{cat} : \mathsf{np}], u_{pn} \lhd u_{np}]$

The view of parsing we follow is one in which this process is seen as a form of inference, i.e., a process which augments an interpretation that can be characterized in logical terms. This is essentially what Sturt and Crocker do, as well; but whereas they characterize the updates to interpretation in an algorithmic fashion, we will characterize these inferences by means of inference rules, as in the 'parsing as deduction' tradition (Lambek, 1958; König, 1989; Shieber *et al.*, 1995) and in inferential approaches to disambiguation (Hobbs *et al.*, 1993; Alshawi, 1992; Lascarides and Asher, 1993). More specifically, we view the rules that infer phrase structure as inference rules (rather than as axioms) as in the 'parsing as natural deduction' approach to syntactic inference (König, 1989); but we interpret these rules in a rather different way, to be discussed in Section §7.

In theories of grammar such as HPSG, categorial grammar, or lexicalized TAGs, all syntactic knowledge is either encoded in the lexicon–e.g., in the form of subcategorization rules– or is specified by a few principles of general validity. In the particular version of lexicalized TAGs adopted by Sturt and Crocker (1996) there are four such operations: Left and Right Attachment, TAG Adjunction, and Tree Lowering. We assume the same set of operations here. They can be encoded in our model by means of general-purpose inference rules reminiscent of those of categorial grammar: for example, Left Attachment, which results in the unification of a subtree with top node A with the leftmost leaf of type A of a second subtree (as shown in Fig. 3), can be expressed in terms of information about utterances as follows:[18]

***Left Attachment*:**

[17]We briefly note here that the properties of the syntactic relation of constituency are preserved by interpreting it in terms of decomposition relations between events, under reasonable assumptions about decomposition; more on this below.

[18]We use the notation $u_1 \trianglelefteq u_2$ to indicate that event u_1 occurs 'immediately before' event u_2; recall also that $u_1 \mathbf{is} u_2$ means that the two discourse entities refer to the same object.

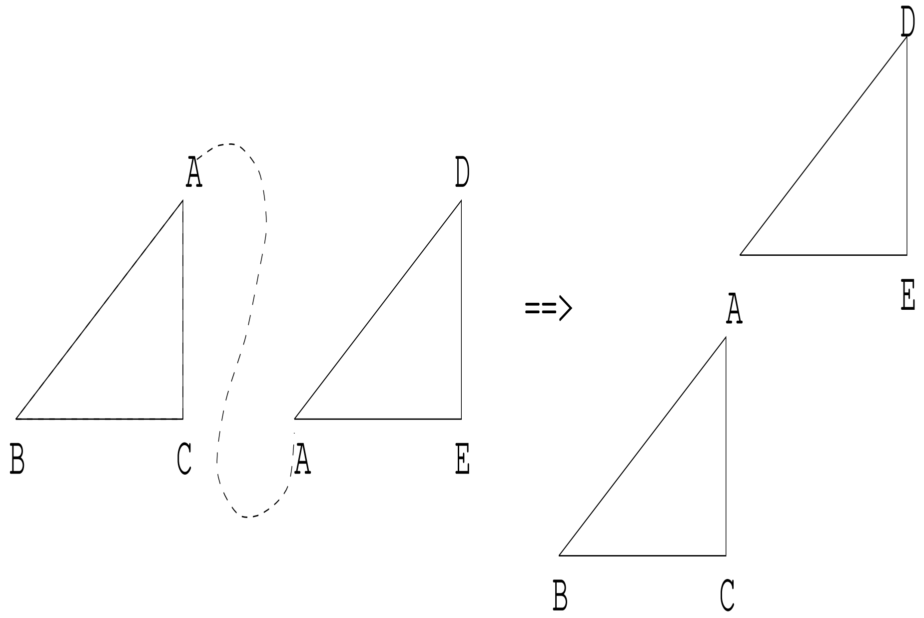


Figure 3: Left Attachment

$$\frac{[|\mathbf{syn}(u_1)\ \mathbf{is}\ \begin{bmatrix}\mathsf{cat}:\mathsf{A}\end{bmatrix}, \mathbf{syn}(u_2)\ \mathbf{is}\ \begin{bmatrix}\mathsf{cat}:\mathsf{A}\end{bmatrix}, \mathbf{syn}(u_3)\ \mathbf{is}\ \begin{bmatrix}\mathsf{cat}:\mathsf{D}\end{bmatrix}, u_2 \lhd u_3, u_1 \unlhd u_2]}{[|u_1\ \mathbf{is}\ u_2]}$$

Semantic Composition

Once the recipient of u_{np} has complete information about the structural relations between this utterance and its subordinated utterances, he/she can derive the conventional meaning of phrasal utterances such as u_{np} from the meaning of lexical items by means of semantic rules that compositionally derive the meaning of an utterance from the meaning of its constituent utterances (i.e., of the subordinated events) just as seen in Section §2.1, except that they now operate in an incremental fashion–i.e., they can apply as soon as the structural relation constituents have been inferred.[19]

[19]Strictly speaking, the interpretation (27) is not yet complete from a syntactic point of view, for the addressee is not guaranteed that he/she has complete information about the decomposition of phrasal utterances. Such guarantee could be given by adding to (27) statements of the form **COMPLETE**(u,u_{np},u_{vp}), defined as follows:

Knowledge about (conventional) meaning composition, just as knowledge about phrase structure, can be formalized in terms of 'inference rules' that specify how the meaning of an utterance is derived from those of its constituents.(Again, we will postpone giving a formal interpretation of these rules until Section §7.) We assume that these rules are distinct from, although depending on, the rules for syntactic composition. For the purposes of the present paper two semantic composition rules will be sufficient. The first, ***Binary Semantic Composition***, encodes type-driven application; the second, ***Unary Semantic Composition***, allows the conventional meaning of a constituent to percolate up when that constituent is unique.

Binary Semantic Composition:
$$\frac{\begin{array}{c} u_1 \lhd u_3, u_2 \lhd u_3 \\ u_1 \rightsquigarrow \varphi_{\langle\alpha,\beta\rangle}, u_2 \rightsquigarrow \psi\alpha, \\ \textbf{COMPLETE}_3(u_3, u_1, u_2) \end{array}}{u_3 \rightsquigarrow \varphi(\psi)}$$

Unary Semantic Composition:
$$\frac{\begin{array}{c} u_1 \lhd u_2, u_1 \rightsquigarrow \varphi, \\ \textbf{COMPLETE}_2(u_2, u_1) \end{array}}{u_2 \rightsquigarrow \varphi}$$

The **Binary Semantic Composition** rule specifies that the conventional meaning of an utterance u_3 whose sole constituents are u_1 and u_2, with meanings φ and ψ, of type $\langle\alpha,\beta\rangle$ and β, respectively, is obtained by applying φ to ψ. Crucially, **Binary Semantic Composition** only applies after both the dominance relations and the meaning of the constituents have been determined; hence, underspecified lexical meanings or underspecified syntactic relations have the effect that the meaning of larger utterances is also underspecified. Similarly, **Unary Semantic Composition** states that the conventional meaning of an utterance u_2 with a single constituent u_1, with meaning φ, is φ.[20]

In the example we have been considering, after parsing has produced (27) **Unary Semantic Composition** can apply, which results in the conven-

$$\forall x, y, z, \textbf{COMPLETE}(x, y, z) =_{def} \forall w w \lhd x \rightarrow w = y \vee w = z$$

which also rely on the assumption that phrasal utterances always decompose in at most two sub-utterances. For simplicity, however, we will omit such statements below.

[20]The binary version of the **COMPLETE** predicate can be defined in the same way as the ternary version.

tional meaning of u_{pn} percolating up to u_{np}, just as in example (4), producing (28):

(28) $K'''' = K'''; [|u_{np} \rightsquigarrow \lambda P[x_{pn}|x_{pn} = k]; P(x_{pn})]$

The Next Utterance

The process of updating the current semantic interpretations continues until no more inferences are possible (let us say until (28) is obtained), and starts again when the next utterance, *croaked*, is perceived. This perception results in K'''' being updated as follows:

(29) $K''''; [u_{iv}|u_{iv}$:**utter**(a,*"croaked"*)]

This new utterance is processed much in the same way, Things continue much in the same way, except that now we have a conflict among two distinct lexical entries, (2) and (3). Let us assume that the lexical entry (2) is chosen for *croaked*; our proposed formalization of the disambiguation process in this case– i.e., how multiple hypotheses are generated and one of them is chosen–is discussed in Section §4. The update to K'''' resulting from accessing lexical entry (2) is shown in (29):

(30) (29); $[|u_{iv} \rightsquigarrow \lambda x[e_m|e_m{:}\mathbf{croak}_f(x), e_m < S], \mathbf{syn}(u_{iv})$ **is** $[\mathsf{cat} : \mathsf{iv}]]$

Again, lexical access is followed by parsing and semantic composition, resulting in the interpretation in (31):

(31) (30); $[u_{vp}|u_{iv} \rightsquigarrow \lambda x[e_2|e_2{:}\mathbf{croak}_f(x), e_2 < S], \mathbf{syn}(u_{iv})$ **is** $\left[\mathsf{cat} : \mathsf{iv}\right]$,
$\mathbf{syn}(u_{vp})$ **is** $\left[\mathsf{cat} : \mathsf{vp}\right], u_{iv} \lhd u_{vp}, u_{vp} \rightsquigarrow \lambda x[e_2|e_2{:}\mathbf{croak}_f(x), e_2 < S]]$

The last update to the interpretation involves hypothesizing that u_{np} and u_{vp} are in turn subconstituents of a larger u_s of type S and assigning it a conventional meaning by means of **Binary Semantic Composition**. The interpretation resulting from these last steps is as follows:

(32) (31b); $[u_s|u_{np} \lhd u_s, u_{vp} \lhd u_s, \mathbf{syn}(u_s)$ **is** $[\mathsf{cat} : \mathsf{s}], u_s \rightsquigarrow [x_{pn}, e_2|e_2{:}\mathbf{croak}_f(x_{pn}), e_2 < S]]$

The reader will have noticed that underspecified interpretations did not play any significant role in this example; this was intentional. In fact, we even adopted a simpler view of syntactic processing than the one proposed by Sturt and Crocker, so that we could postpone dealing with syntactic underspecification. We do this in Section §3.4.[21]

Utterance-Based Notation vs. Sign-Based Notation

In the rest of this paper, we will generally put interpretations such as (26), (27), (30), or (32) in a more readable format by representing the information about phrasal utterances such as u_{np}, which in utterance-based notation would be indicate as:

$$\mathbf{syn}(u_{np}) = \begin{bmatrix} \mathsf{cat} : \mathsf{np} \end{bmatrix} \wedge u_{np} \leadsto \varphi$$

in feature-based notation, as follows:

- $\begin{bmatrix} \mathsf{syn} : \begin{bmatrix} \mathsf{cat} : \mathsf{np} \end{bmatrix} \\ \mathsf{sem} : \varphi \end{bmatrix}_{u_{np}}$

and by representing subordination among utterance events as tree dominance– i.e., representing the interpretation in (32) in a format similar to (4).

Our use of a sign-based notation is also in part meant to make more apparent the similarities between the idea that interpretation involves gathering information about 'utterances' and the idea that interpretation involves gathering information about 'signs'. As a result of this similarity, theories of lexical and syntactic competence developed in a theory such as HPSG can be turned into theories of the contribution of lexical access and parsing to utterance interpretation in the present framework. [22]

[21]The solution to the paradox of incremental interpretation proposed in (Chater *et al.*, 1995) aims to achieve a stronger form of update to the present interpretation than the one we have just discussed, one that they call INCREMENTAL UNDERSTANDING. We are not convinced that incremental understanding in the sense of Chater *et al.* actually takes place; we will return to this topic in Section §4.

[22]The main difference between the utterance-based and the sign-based view of interpretation is that a characterization in terms of utterances makes both the 'building blocks' of an interpretation (the utterances) and the structural relations between them more explicit; thinking in these terms, instead of in terms of a single sign, results in a cleaner formalization of those states of interpretation in which not all structural information has been identified. Another, less important difference is that we do use unification for building interpretations, but lambda-application together with equality among discourse entities (the **is** operator).

Another motivation is that we wanted to have a clear connection between our proposal and the assumptions of mainstream semantic theory. We feel that even though semantics for sign-based notations have been specified (Carpenter, 1992), the connection between this semantics and the assumptions in formal semantics is often unclear; and we believe that our proposal is a plausible way of relating signs to the traditional ontological assumptions of semantic theory.

3.3 A Note on Modularity

In discussing the example above, we talked about 'lexical access', 'parsing', and 'semantic composition' as if these were separate modules, operating in sequence. This was done for expository purposes only. The only modularity assumption that we are going to make (for reasons to be discussed in Section §4.3) is that there is a distinction between a 'language module' and a 'pragmatic' or 'commonsense reasoning' module; we assume that during utterance processing the language module generates hypotheses as discussed in Section §4, which the commonsense reasoning component then accepts or rejects, and possibly ranks. (See also (Poesio, 1994; Lascarides *et al.*, 1996b).) In generating these hypotheses the language module may use certain types of semantic information, typically lexical in nature (e.g., information about selectional restrictions and/or semantic relations between lexical items), but does access general reasoning; however, the interaction may take place at a fairly fine granularity level–e.g., hypotheses about the referent of an anaphoric expression may be tested before the rest of the sentence has been processed, as in (10).

We will not, however, assume any further subdivisions within the language module. Our theory of interpretation is formulated in terms of 'update rules' (introduced in Section §4) that operate on the current interpretation and augment it with new information. Some of these rules are going to operate on the basis of information about the occurrence of an utterance only; we will call them 'lexical access rules', and talk of a 'lexical module'. Other rules operate on the basis of information about the syntactic category of certain utterances, and generate hypotheses about the structural relation of these utterances with other utterances; we call these rules 'parsing rules', and talk of a 'parsing module'. For our purposes, however – specifically, for the discussion in §3.4–and given the formalization of these rules in Section §4, it doesn't matter if all the lexical access rules operate before all the parsing rules, or if these two types of rules are all active together and their temporal sequence depends only on certain update rules depending on the output of other update rules; we can still talk of a (possibly, virtual) 'lexical module' anyway.

3.4 Underspecification

What is, then, underspecification? Or, to be more precise, what does it mean to claim that at some stages of processing, the recipient of an utterance is entertaining an 'underspecified interpretation'?

The characterization above of the semantic interpretations produced by the language processor allows us to give a very simple account of what it means for an interpretation to be underspecified. When one starts trying to do this, however, it turns out that there are at least two distinct senses in which we can say that a module of the language processing faculty produces a partial interpretation; indeed, many of the discussions about underspecification in the computational semantics literature have been concerned with determining the type of underspecified interpretations that occur when interpreting utterances exhibiting a given type of ambiguity. We will discuss these two types of underspecification in turn.

H-underspecification

We begin by defining what we mean when we say that the language processor has 'complete' information about an utterance. Let us assume that there is a special syntactic category `utt` for 'syntactically complete' utterances. Then, we can say that

Definition 1 *The language processor has* COMPLETE INFORMATION *about an utterance u iff the interpretation it entertains specifies the values of $\rightsquigarrow$ and* **syn** *for u and, in case* **cat**(**syn**(*u*)) $\neq$ `utt`*, the value of* $\lhd$.

Since an interpretation I is a Compositional DRT theory, we can also say, more formally, that I is complete with respect to u iff there are objects m, s and u' such that:

$$I \models u \rightsquigarrow m \wedge \mathbf{syn}(u) = s \wedge u' \lhd d.$$

Not all of the interpretations discussed in the previous section are 'complete' in the sense of Definition 1, since none of the modules of the language processor discussed in §3.2 provides complete information about an utterance. Neither the initial interpretation of u_{pn} in (24), K', nor K'' in (26) are complete with respect to u_{pn}; only K''' is. We are going to use the term H-UNDERSPECIFIED to refer to interpretations which are 'partial' or 'incomplete' in this sense:[23]

Definition 2 *An interpretation is* H-UNDERSPECIFIED *with respect to utterance u if it is not complete, i.e., if it does not specify the value of either $\rightsquigarrow$,* **syn**(*u*)*,* **phon**(*u*)*, or* $\lhd$.

[23]The term h-underspecified is used because this is the type of underspecification one finds in cases of homonymy, as discussed below.

We can also introduce a notion of h-underspecification relativized to a given function *f*, where *f* is one of $\rightsquigarrow$, **syn**(u), **phon**(u), or $\lhd$, and say that an interpretation I is h-underspecified with respect to utterance u and function *f* iff there is no value v in *Dom(f)* such that $I \models f(u) = v$.

Clearly, just saying that language processing involves interpretations that are h-underspecified in the sense of Definition 2 is not a very interesting claim, since it's true by definition. As the recipient of an utterance of *Kermit croaked* has to recover all components of the interpretation in (32) on the basis only of the spoken or written signal (in (24)) and his knowledge of grammar and contextual processing, the input to the language processor must be h-underspecified; and all modules involved in utterance processing—lexical access, parsing, anaphoric processing, scope assignment, and so forth—must map h-underspecified interpretations into h-underspecified interpretations. If a module always outputted complete interpretations, there would be nothing left for the other modules to do.

It's also possible, however, to make much more interesting claims about the role of underspecification. In order to do this, we need to specify the TASK of each module of the human language processor in terms of the information that that module can be expected to provide. So, for example, we could define the task of the lexicon with respect to utterance u as that of specifying values for $\rightsquigarrow$ and **syn**, whereas the task of the parser is to specify values of $\lhd$. (I.e., we could say that the task of the lexicon is to produce an interpretation I such that there are values m and s for which $I \models u \rightsquigarrow m \wedge u$ **syn** s.) We can then define a MODULE-COMPLETE INTERPRETATION of an utterance as follows:

Definition 3 *An interpretation of utterance u is* COMPLETE WITH RESPECT TO MODULE M *(or* M-COMPLETE, *for short) iff it specifies all of the information that is part of the task of M to provide.*

Until now, we have only seen examples of interpretation processes outputting M-complete interpretations: i.e., we have seen cases of lexical access resulting in lexicon-complete information about the particular lexical item, and of parsing resulting in parser-complete information about a given utterance. Things get more interesting when we consider the possibility that certain modules may not always produce M-complete interpretations: for example, that the lexicon does not always specify values for $\rightsquigarrow$ and **syn**.

The evidence discussed in Section §2 shows that in some cases we have h-underspecified interpretations in input and M-complete (although possibly still h-underspecified) interpretations in output. For example, both the psychological evidence reviewed in that section, and linguistic evidence from the so-called *ambiguity tests* (Zwicky and Sadock, 1975), suggest that in the case of homonyms[24] distinct and lexicon-complete lexical entries are accessed, so that interpretations like (30) are recovered from (29) directly, without going through any 'intermediate' interpretations (more on this in §4). However, we will see in Section §5 evidence that the lexical entry of pronouns is h-underspecified; indeed, this is one of the main claims of this paper.

The characterization of h-underspecification we have just given is general enough to cover not only the possibility of partiality in the input and output of the lexicon and of modules concerned with semantic interpretation, but also of syntactic h-underspecification. The syntactically underspecified interpretations proposed by the theories of parsing based on d-theory discussed in Section §2.4 can be seen as cases of h-underspecification, as well, except that in these cases the interpretation specifies constraints on the ways the missing value (of $\lhd$) can be resolved. Those interpretations can be reformulated in terms of Compositional DRT by interpreting immediate dominance relations between nodes of a tree as direct subordination relations among utterance events, and indirect dominance in terms of the transitive closure of $\lhd$, that we may indicate as $\lhd^*$. The information expressed by the interpretation in Figure 2 could then be expressed as in (33).

(33)
$$
\begin{aligned}
& u_{pro}\textbf{:utter}(a,\text{“}I\text{”}) \land \textbf{syn}(u_{pro}) = \begin{bmatrix}\mathsf{cat : pro}\end{bmatrix} \land u_{pro} \leadsto a \\
& \land\ \textbf{syn}(u_{np_1}) = \begin{bmatrix}\mathsf{cat : np}\end{bmatrix} \land u_{pro} \lhd u_{np_1} \land u_{np_1} \leadsto a \\
& \land\ u_{tv}\textbf{:utter}(a,\text{“}saw\text{”}) \land \textbf{syn}(u_{tv}) = \begin{bmatrix}\mathsf{cat : tv}\end{bmatrix} \land u_{tv} \leadsto \textbf{saw} \\
& \land\ \textbf{syn}(u_{vp}) = \begin{bmatrix}\mathsf{cat : vp}\end{bmatrix} \land u_{tv} \lhd u_{vp} \land \\
& \textbf{syn}(u_s) = \begin{bmatrix}\mathsf{cat : s}\end{bmatrix} \land u_{np_1} \lhd u_s \land u_{vp} \lhd u_s \land \\
& \textbf{syn}(u_{np_2}) = \begin{bmatrix}\mathsf{cat : np}\end{bmatrix} \land u_{np_2}\textbf{:utter}(a,\text{“}a\ man\text{”}) \land u_{np_2} \lhd^* u_{vp} \land \\
& u_p\textbf{:utter}(a,\text{“}with\text{”}) \land \textbf{syn}(u_p) = \begin{bmatrix}\mathsf{cat : p}\end{bmatrix} \land u_p \leadsto \textbf{with} \land
\end{aligned}
$$

[24]We recall that these are the 'two-meanings' words of Frazier and Rayner–words associated to lexical entries of different syntactic categories (such as *can*) or words such as *bank* or *stock* associated with more than one lexical entry of the same syntactic category.

$$\mathbf{syn}(u_{np_3}) = \begin{bmatrix}\mathsf{cat} : \mathsf{np}\end{bmatrix} \wedge u_{np_3}\text{:}\mathbf{utter}(a,\text{“}\textit{a telescope}\text{”}) \wedge$$

$$\mathbf{syn}(u_{pp}) = \begin{bmatrix}\mathsf{cat} : \mathsf{pp}\end{bmatrix} \wedge u_p \lhd u_{pp} \wedge u_{np_3} \lhd u_{pp} \wedge u_{pp} \lhd^* u_{vp}$$

Our notion of interpretation can therefore be used to encode the syntactic information in the interpretations proposed in (Sturt and Crocker, 1996), as well.

There have been suggestions that the parser may output h-underspecified interpretations that do not even specify such constraints: for example, in the case of PP attachment (Marcus *et al.*, 1983; Abney, 1991). These would be interpretations which, unlike the interpretation in Figure 2, do not even specify restrictions on the possible attachment points of the PP *with a telescope*. This seems especially likely in the case of spoken input: in the case of conversations like that in (34), for example, it is often assumed that the parser only outputs a series of 'chunks,' which are then assembled by the process concerned with speech act interpretation.

(34) a

Finally, perhaps the most interesting claim about the role of h-underspecified interpretations in utterance interpretation is the hypothesis that the output of the 'linguistic module' as a whole can be h-underspecified. Unsurprisingly, this claim has attracted a lot of attention; in particular, a number of theories of underspecification have been proposed to specify the interpretation of scopally ambiguous sentences (Alshawi and Crouch, 1992; Poesio, 1991, 1996a; Reyle, 1993). Unfortunately, there isn't enough evidence yet about which types of scopal relations are resolved incrementally and which ones are left underspecified; we will therefore postpone discussing scopal underspecification until Section §6.

P-underspecification

There is a second sense in which we can talk of an interpretation being 'partial'. This happens when a module outputs an interpretation that does specify a value v for a given attribute f of an utterance, but v is not the actual value of f; rather, it subsumes this actual value, that has to be determined later–often at the end of the sentence–by restricting this initial interpretation. We will use the term P-UNDERSPECIFICATION for this type of underspecification, that can be defined more precisely in terms of disjunction as follows:[25]

Definition 4 *An interpretation I is* P-UNDERSPECIFIED WITH RESPECT TO UTTERANCE *u* AND FUNCTION *f (where f if one of* $\rightsquigarrow$, **syn**(*u*), **phon**(*u*), *or* $\lhd$) *iff there are values* $v_1, \ldots v_n$ *such that*

$$I \models f(u) = v \wedge (v = v_1 \vee \ldots \vee v = v_n).$$

The initial interpretations assigned to polysemous words (the 'multiple senses' cases of Frazier and Rayner (1990)) are cases of p-underspecification.[26] According to Pinkal (1995); Cruse (1995); Copestake and Briscoe (1995), among others, expressions such as *head* or *newspaper* that have distinct, but related, senses[27] have a single lexical entry, but this entry does not completely determine the meaning of an utterance of that expression: it only specifies a meaning that subsumes the distinct senses. For example, the interpretation of *newspaper* specified by its lexical entry can be characterized as in (35), where $\mathbf{newspaper}_U$ is the hypothesized underspecified predicate that subsumes the predicates associated with the distinct senses of *newspaper*. (Newspaper as a physical object, newspaper as its content, newspaper as an institution, etc.)[28]

(35) u_n:**utter**(*a*,"*newspaper*") $\wedge\ u_n \rightsquigarrow$ *newspaper*$_U$

In other words, this hypothesis amounts to a claim that in addition to word-strings like *Kermit* associated with a single, fully specified lexical entry such as (1), and word-strings like *croaked* (or *bank* or *pitcher*) associated

---

[25] H-underspecification could be seen as the limiting case of p-underspecification in which an interpretation I only supports the inference that one among all values in the range of *f* may be its value – $\exists x f(u) = x$ – but in practice, such constraints are only interesting when they are not vacuous in this sense. See, however, Pinkal (1995) for a theory of ambiguity developed on the basis of the claim that there are no sharp boundaries between p-underspecification and h-underspecification.

[26] Hence the name.

[27] In addition to the head of a human being, we can have the head of a nail, the head of a corporation, etc., all senses that share some aspects of the 'human head' sense.

[28] A word of caution is necessary here. As noted by Pinkal (1995) and Lyons (1995), among others, it is in general very difficult, if not impossible, to identify with absolute certainty the discrete 'senses' of a polysemous word; in fact, it is often difficult even to decide whether we are dealing with a case of homonymy or polysemy (*croaked* is an example of this). To some extent, whether the different interpretations of a string are associated with distinct lexical entries, or not, is a matter of individual differences. As we will see below, our proposal for lexical entries does not draw a sharp distinction between the two types of interpretations.

with two or more fully specified lexical entries such as (2) and (3),[29] there are word-strings such as *newspaper* or *head* associated with a single, but p-underspecified lexical entry. The proposed lexical entry for *newspaper* can be represented in sign-based notation as in (36):[30]

(36) $newspaper_m \rightsquigarrow \begin{bmatrix} \mathsf{phon} : \text{“newspaper''} \\ \mathsf{syn} : \begin{bmatrix} \mathsf{cat} : \mathsf{n} \end{bmatrix} \\ \mathsf{sem} : \lambda x[\ |\mathbf{newspaper}_U(x)] \end{bmatrix}_{u_m}$

When processing sentences such as those in Frazier and Rayner's Set 3 (in (15)), the human processor initially builds an interpretation of the utterance of *newspaper* as in (35). This initial conventional meaning is then combined with the meaning of the rest of the words in the utterance, resulting in an interpretation like (37):

(37) u:**utter**(a,*"Lying in the rain, the newspaper was destroyed."*)
$\wedge\ u \rightsquigarrow [x_n, y_p, e_m, e_q \mid \mathbf{newspaper}_U(x_n), e_m{:}\mathbf{destroyed}(x_n), e_m < S,$
$\mathbf{rain}(y_p), e_m{:}\mathbf{lie\text{-}in}(x_n, y_p)\ ]$

From this initial interpretation, the more specialized reading that x_n is actually a $\mathbf{newspaper_{physobj}}$ is then derived by means of rules that depend on other sorts of information (e.g., on the basis of the selectional restrictions of *lying in the rain*, that presumably requires a physical object as its argument). These inferences, which on the basis of the results of Frazier and Rayner take place at the end of the sentence rather than incrementally, result in an interpretation as in (38).

(38) u:**utter**(a,*"Lying in the rain, the newspaper was destroyed."*)
$\wedge\ u \rightsquigarrow [x_n, y_p, e_m, e_q \mid \mathbf{newspaper_{physobj}}(x_n), e_m{:}\mathbf{destroyed}(x_n), e_m < S,$
$\mathbf{rain}(y_p), e_m{:}\mathbf{lie\text{-}in}(x_n, y_p)\ ]$

A special case of p-underspecified interpretation is the one where the property initially attributed to the object is one which holds for all objects and therefore does not restrict its interpretation at all–i.e., the object is asserted

[29] Of course, these distinct lexical entries may also differ in that they specify different syntactic properties such the syntactic category–as in *can* that can be either a noun or an auxiliary–instead of different semantic properties.

[30] A similar situation holds for strings like *sheep* some of whose syntactic properties–number, in this case–are p-underspecified; several proposals to handle this type of syntactic underspecification have been made in the syntactic literature.

to be a **thing**. Interpretations such as these have been proposed by Hobbs and associates as the initial interpretations in a number of interpretation problems that they claim are solved on the basis of what they call 'local pragmatics' (Hobbs and Martin, 1987; Hobbs *et al.*, 1993), and which include, in addition to lexical disambiguation, anaphoric processing, and determining the relation between nouns in a noun-noun compound.

It has also been claimed that p-underspecified interpretations are derived when disambiguating plurals (Verkuyl and van der Does, 1991), and even that such 'most general' interpretations are obtained at an initial stage of scope disambiguation (Kempson and Cormack, 1981), but this proposal is now generally regarded as implausible (see Section §2.4).

3.5 The Formal Language: A Minimal Extension of Compositional DRT

As said above, most of the information in (25), (31) and (32) can be specified using the formal tools provided in Compositional DRT, and therefore no additional logical machinery will be needed to characterize the interpretation process.[31] To the language of Compositional DRT (specified in Appendix §A) we only add the following:

1. the expression e:$\mathbf{p}(\overline{x})$, used as a shorthand for what Muskens would write as $\mathbf{p}(\overline{x}, e)$;

2. a set of non-logical functions $\leadsto_\alpha$ of type $\langle\epsilon,\alpha\rangle$, one for each type α in Θ. We will generally abbreviate this into $\leadsto$.[32]

[31] Most work on semantic underspecification, including our own previous work, has involved developing special logics of semantic interpretation that differ in radical ways from the logics used in mainstream semantic theory. For example, in (Poesio, 1994, 1996a) we proposed that unambiguous sentences denote a singleton set whose single element is a proposition, whereas an ambiguous sentence like *Kermit croaked* denotes a set consisting of two or more propositions; the interpretation of a sentence remains underspecified as long as it corresponds to a set of propositions of cardinality greater than one. The introduction of meta-disjunctions of this sort necessitates special (partial) logics: e.g., the logics in Alshawi and Crouch (1992); Poesio (1996a) are similar to Kleene's 3-valued logic. (See (Pinkal, 1995) for a discussion of partial logics for ambiguous sentences, and (van Deemter, 1991, 1998) for a discussion of possible options in setting up an ambiguous logic.) We now believe these special logics to be unnecessary for the task of characterizing underspecified interpretations.

[32] It may be argued that it's not very plausible to assume that each utterance can have a meaning of all types. The solution is to assume that the functions $\leadsto_\alpha$ are partial functions depending on the (syntactic) type of the utterance event. We will ignore this issue here.

3. the function $\lhd$ and the relation $\lhd^*$, both of type $\langle\epsilon,\langle\epsilon,t\rangle\rangle$

The only constraint imposed on the functions $\leadsto_\alpha$ is that they are functions:

AX12 $\forall x_\epsilon y_\alpha z_\alpha (x \leadsto_\alpha y \land x \leadsto_\alpha z) \rightarrow y = z$

The other properties of $\leadsto_\alpha$ will be specified by the rules for disambiguation, that we will discuss in the next section. As for $\lhd$, we assume it has the properties typically assumed of relations among nodes of a tree (Marcus *et al.*, 1983), as axiomatized, e.g., in (Backofen *et al.*, 1995), but we expect most of these properties to follow from the intended interpretation of $\lhd$ as a representation of the decomposition relation among events, rather that from a special stipulation. Muskens provides only a bare-bones axiomatization of events and only includes an event-inclusion relation derived from temporal inclusion, but some of the desired properties of $\lhd$ already follow under the assumption that decomposition entails event inclusion.[33]

4 How Semantic Interpretations are Produced: The Case of Lexical Access

In the previous section, we have introduced the formal tools necessary to characterize the fully specified semantic interpretations produced incrementally by the language processor, the p-underspecified interpretations produced by the language processor before reaching a fully specified interpretation (e.g., in the cases studied by Frazier and Rayner), and also the

[33] One possible exception is that in general we do not want to assume that we are dealing with a single tree, i.e., that all utterances are going to be constituents to a single larger utterance. This is because in spoken language, processes such as parsing, reference resolution and semantic interpretation interact with other processes, such as turn-taking (Levinson, 1983), grounding (Traum, 1994), and repairs of the output being produced. As a result, the spoken signal includes a lot of hesitations, restarts, and other so-called 'disfluencies'. In addition, a speaker may interrupt his own utterance in order to ask a clarification question.

The crucial point here is that a lot of these 'disfluencies' are actually utterances with a well-determined purpose; it's just that this purpose is not to contribute to the conventional meaning of a single utterance. Instead, the goal of these utterances is to ensure that some conversational subtask or the other gets accomplished, or to acquire some information necessary before going on completing the current utterance. We cannot simply assume a model of utterance processing in which the language processor tries to arrive at an interpretation of the current utterance discards utterances such as *uh* or *okay*, or clarification questions: they all play a role, although not necessarily in specifying the conventional meaning of the initial utterance.

h-underspecified interpretations entertained by the processor at intermediate stages of processing, and (according to Garrod and Sanford) at the end of processing in the case of certain pronouns. However, we did not say much about the 'update rules' producing these interpretations; also, we did not say how to characterize a state in which the language processor finds itself frequently, on the basis of what see in Section §2–namely, the case in which more than one (fully specified) interpretation is entertained in parallel, just briefly before being pruned (as in the cases studied by Swinney) or until the end of the sentence (as in the cases studied by Corbett and Chang) or even later (as in the examples discussed in Section §2.5). The machinery introduced in this section (using lexical access as an illustration) will fill both gaps in the theory, which we then apply to pronoun interpretation in Section §5. Our main claim in this section is going to be that the formal characterization of the multiple hypothesis scenarios discussed in Section §2 falls out naturally from the most natural (and most widely accepted) way of formalizing the rules for updating semantic interpretations.

4.1 Semantic Processing as Defeasible Reasoning

All the psychological evidence we are aware of suggests that human reasoning is DEFEASIBLE, in the sense that humans jump to conclusions in the absence of sufficient information, and therefore their conclusions may have to be retracted. The best known example of commonsense reasoning is concluding that *Tweety flies* if we know that *Tweety is a bird*; this may seem a natural conclusion to draw, yet may have to be retracted if we learn later that Tweety is a penguin. Many experiments on human concept formation and categorization show that humans draw inferences of this type (Rosch, 1975). Clark (1977) showed that subjects draw defeasible 'bridging' inferences in order to interpret anaphoric expressions, while Johnson *et al.* (1973) showed that their subjects performed defeasible causal inferences. (See, e.g., (Oaksford and Chater, 1991) for a discussion.)

A crucial property of defeasible reasoning is that jumping to conclusions in the absence of complete evidence means that more than one hypothesis may be consistent with what we already know. In any context in which more than one rule will be applicable, we may get conflicting predictions. For example, we may have a defeasible rule that says that Republicans are not pacifists, and another that says that Quakers are pacifists; if we know that Nixon is both a Republican and a Quaker, we get a conflict, since we may conclude both that he is a pacifist and that he is not a pacifist.[34]

The evidence discussed in §2 indicates that utterance disambiguation is a defeasible reasoning process, just like other aspects of human reasoning. This is suggested, first of all, by its incremental nature: the fact that humans resolve ambiguities so quickly indicates that they jump to conclusions, rather than waiting for disambiguating information. The second indication that disambiguation is defeasible in character is the fact that multiple interpretations are generated in parallel, and that more than one interpretation may still be perceived even at the end of the sentence.

These observations are not new, of course; on the contrary, most theories of utterance processing proposed in computational linguistics are formulated in terms of defeasible rules, even though only a few of these theories are explicitly formulated in terms of a non-monotonic or statistically-based inference formalism. The theories formulated (implicitly or explicitly) in defeasible terms include the theories of lexical access proposed by (Hirst, 1987; Lascarides *et al.*, 1996b; Schütze, 1997); most current theories of parsing, whether based on principles such as Minimal Attachment (Frazier, 1987) or on statistical preferences (MacDonald *et al.*, 1994; Stolcke, 1994; Jurafsky, 1996); theories of anaphoric interpretation such as Centering Theory (Grosz *et al.*, 1995), that involve defeasible inferences both in determining which entity is currently most salient, and in concluding that the most salient nominal entity is the most likely antecedent of an anaphoric expression; and theories of tense interpretation such as (Lascarides and Asher, 1993; Kameyama *et al.*, 1993). In fact, the claim that utterance processing is a form of defeasible reasoning–more specifically, a form of abductive reasoning–is a central claim of the theories of discourse interpretation developed at SRI Palo Alto and SRI Cambridge (Hobbs *et al.*, 1993; Alshawi, 1992). In proposing a formalization of disambiguation in terms of defeasible reasoning we are therefore following in the path of a number of other researchers, even though only a few of the more formalized proposals have been explicitly set up to account for psychological evidence, as we try to do here, and no other researchers have tried to use the properties of defeasible reasoning to clarify the notion of ambiguity.

Although the claim that disambiguation is a defeasible inference process is largely accepted, disagreement begins as soon as one attempts to be more precise about the nature of the defeasible formalism. A particularly heated debate is that between accounts of defeasible reasoning based

[34]This kind of conflict is known as a NIXON DIAMOND, for the obvious reasons (Lascarides and Asher, 1993).

on statistics and accounts based on nonmonotonic logics (Oaksford and Chater, 1991). We would like to avoid committing ourselves to a particular position on this point. Our position is that, in all likelihood, disambiguation is based on statistical information (Hwang and Schubert, 1993; MacDonald *et al.*, 1994; Jurafsky, 1996; Schütze, 1997); a statistical account of disambiguation comes with a natural notion of priority among rules (see below) and of evidence combination, and may also account for how disambiguation rules are acquired. Unfortunately, however, nobody knows yet how to merge a probabilistic logic with the types of logics proposed in formal semantics to account for linguistic evidence.

We have therefore formulated our theory of disambiguation in terms of what is perhaps the best known nonmonotonic logic, DEFAULT LOGIC (Reiter, 1980). This logic is well-known, and allows us to account for the properties of the disambiguation process in which we are interested; in addition, it is rather simple (which makes it easy for us to show how the properties of disambiguation seen above are derived) and it is easy to modify to make it into an extension of a logic such as Compositional DRT, since it is developed proof-theoretically instead of semantically (we return to this below). The considerations above suggest, however, that the formalization proposed below should be seen as a first approximation, to be revised when more appropriate formalisms become available. Fortunately the connection between nonmonotonic logics and formalisms for statistical reasoning is becoming clearer (Geffner and Pearl, 1990; Bacchus *et al.*, 1993; Poole, 1993); this makes it reasonable to expect a greater convergence between the logical theories used in formal semantics and statistically-based theories of reasoning in the future.[35]

[35]The fact that the formalization we present here is in terms of inferences in a logic also needs some justification, since many theories of disambiguation give what we may call an 'algorithmic' account of disambiguation: i.e., they how a given disambiguation is obtained in terms of algorithms (Brennan *et al.*, 1987; Hobbs and Shieber, 1987; Hwang and Schubert, 1992; Jurafsky, 1996). Our main reason for using a logic is that we want our account of disambiguation to give us a way to fill in what, given what we have said so far, may have seemed like a strange gap in a theory of ambiguity processing. Under the semantics of underspecified interpretations that we have proposed, expressions that we may want to call 'ambiguous', such as the word *croaked*, are assigned partial interpretations. Yet, there is a strong intuition that ambiguity has something to do with multiplicity of meaning: this is the definition one finds in dictionaries, an even formal characterizations of ambiguity such as (Pinkal, 1995) are specified in terms of the 'precisifications' of the meaning of the expression. Where is this intuition about the meaning of ambiguous expressions captured in the theory we have presented? We are going to propose that the multiple meanings of an utterance correspond to alternative ways of completing an interpretation in the logic of disambiguation that we will present. The fact that the theory of disambiguation ends up

A second point that we should mention is that the computational properties of most nonmonotonic logics, including Default Logic, make it unlikely that they can be viewed as the ultimate account of human reasoning (Oaksford and Chater, 1991). Because Default Logic is built on top of the notion of provability in another logic, it isn't even semi-decidable. Clearly, a plausible foundation for a theory of human reasoning would have to satisfy complexity requirements, as well. Nonmonotonic logics with better complexity properties have been proposed, among others, in Selman and Kautz (1990); Lascarides *et al.* (1996a); however, our only concern here is to provide a characterization of the properties of the disambiguation process discussed in Section §2, and therefore we consider using Default Logic an acceptable simplification. In addition, we believe that the complexity problem has to be addressed not by expecting a single type of logic to account for all types of disambiguation , but by providing more specific formalizations of each disambiguation process in terms of weaker formalisms that do not suffer from the problems just mentioned, but have the very general properties assumed here. This means, however, that our readers should not expect from this paper predictions based on the computational properties of the formalization of semantic processing we propose.

4.2 Default Logic

This section is a short reminder of the main characteristics of Default Logic. The reader interested in more details should consult Reiter's paper or a textbook such as (Brewka, 1991).

Default Logic is a generalization of first order logic that allows for DEFAULT INFERENCE RULES in addition to standard inference rules. A default inference rule takes the form $\alpha(X) : \beta_1(X) \ldots \beta_n(X)/\gamma(X)$, where $\alpha(X)$ is the PRECONDITION, $\gamma(X)$ is the CONSEQUENT and $\beta_1(X) \ldots \beta_n(X)$ is the set of JUSTIFICATIONS. Such a rule should be read: if α(X) is the case, and if it is consistent to assume that β(X), conclude that γ(X). For example, the default that *Quakers are Pacifists* is formalized by the default inference rule **QUAKER-PACIFIST**, shown below,

$$\frac{\textbf{quaker}(X) \ : \textbf{pacifist}(X)}{\textbf{pacifist}(X)}\textbf{QUAKER-PACIFIST}$$

providing part of the characterization of ambiguous expressions is, in fact, an important reason for choosing to formalize disambiguation in logical terms.

whose interpretation is as follows: if X is a Quaker, and if it is consistent to assume that X is a Pacifist, then we can assume that X is indeed a Pacifist. This particular rule, as all the rules that we will use in the paper, is an OPEN DEFAULT containing SCHEMA VARIABLES such as X. These defaults behave like inference schemas, in the sense that they denote the infinity of CLOSED DEFAULTS obtained by replacing the schema variables with constants of the language. We will indicate schema variables with uppercase letters.

A DEFAULT THEORY $\langle$D,W$\rangle$ consists of a set of formulas W paired with a set of defaults D. An EXTENSION of a default theory $\langle$D,W$\rangle$ is the deductive closure of the theory under a subset of the default inference rules in D: intuitively, as many as can be added while maintaining consistency. The extensions of a default theory are the fixed points of the operator Γ, defined as follows:

- Let Δ = (D,W) be a closed default theory, so that every default of D has the form (α:β_1,..., β_m/w) where α,β_1,...,β_m,w are all closed wffs of L. For any set of closed wffs S $\subseteq$L let Γ(S) be the smallest set satisfying the following three properties:

  D1 W $\subseteq\Gamma$(S)

  D2 Th(Γ(S)) = Γ(S)

  D3 If (α:β_1,..., β_m/w) $\in$ D and $\alpha\in$ S and $\neg\beta_1, \ldots, \neg\beta_m \neg \in$ S then w $\in \Gamma$(S).

Extensions are the formal counterpart of the informal notion of 'hypothesis'; by talking in terms of generating extensions we can abstract away as much as possible from the details of particular processing schemes (whether hypotheses are generated serially or in parallel, for example). Consider the default theory consisting of only one fact— that Nixon is a quaker— and only one default inference rule, the rule **QUAKER-PACIFIST** shown above:

W = {**quaker**(n)}

$$D = \frac{\mathbf{quaker}(X)\ :\ \mathbf{pacifist}(X)}{\mathbf{pacifist}(X)}\ \textbf{QUAKER-PACIFIST}$$

Under the definition of extension given by Reiter, and assuming the notion of provability of first order logic, this theory has only one extension, which contains the fact that Nixon is a pacifist.

Because extensions have to be consistent, but defaults need not be, we may get more than one extension for a theory. Assume we have a default theory that includes, in addition to the default that Quakers are pacifists, a second default, that Republicans are not pacifists:

$$\frac{\textbf{republican}(X)\ :\neg\textbf{pacifist}(X)}{\neg\textbf{pacifist}(X)}\textbf{REPUBLICAN-WARMONGER}$$

assume further that our default theory includes the fact that Nixon is a Republican. Then, the default theory

$$\langle\{\textbf{QUAKER-PACIFIST},\ \textbf{REPUBLICAN-WARMONGER}\}\ ,\ \{\textbf{quaker}(\textit{nixon}),\textbf{republican}(\textit{nixon})\}\rangle$$

will have two extensions, because both defaults are applicable, but an extension cannot include both **pacifist**(x) and $\neg$**pacifist**(x); hence, any extension resulting from applying the default that Quakers are pacifists cannot include the results of applying the other default. The first extension will include the fact that Nixon is a pacifist; the second extension the fact that Nixon is not a pacifist.

This possibility of conflicts between extensions is usually thought of as a problem with defeasible reasoning, but, in fact, we have seen above that humans display precisely this kind of behaviour when interpreting utterances, and writers exploit the reader's ability to obtain conflicting interpretations to achieve a rhetorical effect, as in the case of puns and *double-entendres*. The results of psychological work on lexical ambiguity also indicate that all senses of an ambiguous word are activated (at least in non-primed contexts) (Swinney, 1979; Seidenberg *et al.*, 1982). The same holds for pronoun interpretation (Corbett and Chang, 1983; Stevenson and Vitkovitch, 1986; Gernsbacher, 1989).

One of the reasons for choosing Default Logic as a formalization of defeasible reasoning is that it provides a way of adding defeasible inference rules on top of any logic, without complicating its semantics. Reiter himself assumed first order logic as his underlying logic, but any other logic with a sound and complete definition of consequence will do; thus, the only change we need to do in order to use default inference rules together with Muskens' formalism is to replace the notion of consequence of first order logic with that of Compositional DRT(see the Appendix). The effect of the default inference rules discussed in the rest of the paper will be to concatenate some new information to the DRS characterizing the current interpretation, as in the model discussed in Section §3.2. Some of these rules would take the form of open defaults such as:

$$\frac{\mathrm{K(I)(J)} : \mathrm{[K;K'](I)(L)}}{\mathrm{[K;K'](I)(L)}}\mathbf{D}$$

where I, J and L are open variables ranging over assignments, L is an extension of J, and K and K' are variables ranging over DRSs. To simplify matters, we will force the notation a bit, and write instead:

$$\frac{\mathrm{K} : \mathrm{K'}}{\mathrm{K'}}\mathbf{D}$$

(i.e., both the antecedents and the conclusion of the inference rule will be DRSs).

4.3 Equally Frequent Words in a Non-Primed Context

To illustrate the role of default inference rules in a theory of semantic processing, we will use the case of underspecification elimination studied by Frazier and Rayner: lexical access.

Multiple Meanings

We will begin with the case most typically studied in psychological experiments, that of words with multiple fully disambiguated lexical entries (i.e., the words that Frazier and Rayner characterized as having 'multiple meanings'). As discussed in Section §2.2, the basic finding about lexical access for these words is that in a non-primed context, all equally frequent lexical entries are activated (Swinney, 1979). This finding can be explained if we assume that lexical access is a form of defeasible reasoning involving default inference rules that allow us to jump to conclusions about the interpretation of utterance events such as u_{iv} in (29). For example, the two lexical entries for *croaked* seen in Section §3 ((2) for the sense of 'uttering the sound that frogs utter,' represented by the property $\mathbf{croak}_f$; (3) for the sense 'to die', represented by the property $\mathbf{croak}_d$) can be formalized by the following default inference rules, in which we have used a feature-based notation to characterize the facts about the utterance that 'trigger' the rules (we remind the reader that this feature notation is really only a shorthand for our utterance-based notation, and that according to the mapping specified in footnote 15, those feature structures really correspond to DRSs):[36]

$$\frac{\begin{bmatrix}\text{phon : ``croaked''}\end{bmatrix}_U \quad : \quad \begin{bmatrix}\text{phon : ``croaked''}\\ \text{syn : }\begin{bmatrix}\text{cat : iv}\end{bmatrix}\\ \text{sem : }\lambda x[e_m|e_m{:}\mathbf{croak}_f(x), e_m < S]\end{bmatrix}_U}{\begin{bmatrix}\text{phon : ``croaked''}\\ \text{syn : }\begin{bmatrix}\text{cat : iv}\end{bmatrix}\\ \text{sem : }\lambda x[e_m|e_m{:}\mathbf{croak}_f(x), e_m < S]\end{bmatrix}_U}\ \textbf{LEX-CROAK}_f$$

$$\frac{\begin{bmatrix}\text{phon : ``croaked''}\end{bmatrix}_U \quad : \quad \begin{bmatrix}\text{phon : ``croaked''}\\ \text{syn : }\begin{bmatrix}\text{cat : iv}\end{bmatrix}\\ \text{sem : }\lambda x[e_m|e_m{:}\mathbf{croak}_d(x), e_m < S]\end{bmatrix}_U}{\begin{bmatrix}\text{phon : ``croaked''}\\ \text{syn : }\begin{bmatrix}\text{cat : iv}\end{bmatrix}\\ \text{sem : }\lambda x[e_m|e_m{:}\mathbf{croak}_d(x), e_m < S]\end{bmatrix}_U}\ \textbf{LEX-CROAK}_d$$

A listener whose lexical competence about *croaked* is characterized by the two defaults above will obtain two interpretations when the word is perceived. These interpretations correspond to the extensions of the following default theory:

(39) $<\{\textbf{LEX-CROAK}_d, \textbf{LEX-CROAK}_f\}, \left\{ \begin{bmatrix}\text{phon : ``croaked''}\end{bmatrix}_{u_{iv}} \right\} >$

(where, again, the sign encodes the DRS resulting from the perception of the utterance, as discussed in Section §3). The two extensions of this theory

[36] These rules formalize a view of lexical access akin to that proposed in so-called BOTTOM-UP models (see (Jurafsky, 1996, pg.150-151) for discussion). According to more recent theories of lexical access, humans do not wait until the whole string of phonemes constituting the word has been encountered before retrieving a lexical entry; such models assume that lexical access begins at the sub-word level by retrieving all lexical entries associated with the prefix encountered until then (as in the Cohort model, (Marslen-Wilson, 1984)) or all those associated with the first stressed syllable (as in the MSS model (Cutler and Norris, 1988)). As mentioned above, in this paper we will not be concerned with sub-word disambiguation; such models could however be easily formalized in terms of default rules activated when sub-word utterances are recorded, and by assuming a more fine-grained view of hypothesis generation and pruning–say, the phoneme level.

include the following facts:

$$\textbf{Extension 1:} \quad \begin{bmatrix} \mathsf{phon} : \text{“croaked”} \\ \mathsf{syn} : \begin{bmatrix} \mathsf{cat} : \mathsf{iv} \end{bmatrix} \\ \mathsf{sem} : \lambda x[e_m | e_m\text{:}\mathbf{croak}_f(x), e_m < S] \end{bmatrix}_{u_{iv}}$$

$$\textbf{Extension 2:} \quad \begin{bmatrix} \mathsf{phon} : \text{“croaked”} \\ \mathsf{syn} : \begin{bmatrix} \mathsf{cat} : \mathsf{iv} \end{bmatrix} \\ \mathsf{sem} : \lambda x[e_m | e_m\text{:}\mathbf{croak}_d(x), e_m < S] \end{bmatrix}_{u_{iv}}$$

(The reason why we get two extensions is that two distinct values are assigned as the value of $\leadsto$ for u_{iv}.)

This example should make it clear why we claim that formalisms such as Default Logic are a natural formalization of a parallel theory of language processing (e.g., of the model of lexical access proposed by Swinney): this is because the process of generating multiple hypotheses is naturally modelled in terms of extension generation, in the sense that we can formally characterize the state of the language processor after encountering the word *croaked* in terms of the extensions of the default theory (39).

So, if we assume that the rules that cause the updates of the semantic interpretation seen in Section §3.2 are default inference rules, we automatically get a formal characterization of the states of the language processor that we couldn't characterize using only the formal machinery introduced in Section §3. The incremental nature of lexical access can also be accounted for by assuming that the process of extension generation starts after each word, as in the example in Section §3.2. So, what is still missing to have a complete account of Swinney's results 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; we will get to this shortly.

In the case of words with semantically fully specified lexical entries, factors other than semantic underspecification ensure that disambiguation occurs quickly and effectively: the fact that interpretation takes place incrementally, that there are only a limited number of lexical entries, and that pruning of alternative hypotheses takes place quickly. (We assume that the garden path effects discussed by Frazier and Rayner can be explained in terms of is parallel incremental interpretation with pruning.) Yet even in this case, the concept of semantic h-underspecification plays a role, if only to characterize the task of lexical access (i.e., what kind of information gets added to an interpretation as the result of lexical access). In other words, a theory of the interpretations entertained by the language processor in which h-underspecified interpretations can be characterized affords, even in cases as simple as the one we have been discussing, a cleaner formalization than prior attempts at formalizing lexical disambiguation that either did not rely on the notion of underspecification at all, and therefore would have to model lexical access by extra-logical means, or only allowed for p-underspecification, as in the case of Hobbs' proposals.

Multiple Senses

Let us now consider the second case discussed by Frazier and Rayner: 'multiple senses' words like *newspaper*. In Section §3.4, we proposed that these words have a single lexical entry with a p-underspecified conventional meaning. In terms of the present formalization, this means that only one default inference rule is activated by the occurrence of an utterance of the word *newspaper*; this rule, which encodes the lexical entry in (36), assigns to the word a single p-underspecified interpretation. Such a rule might look like **LEX-NEWSPAPER**:

$$\frac{\begin{bmatrix}\mathsf{phon} : \text{``newspaper''}\end{bmatrix}_U \quad : \quad \begin{bmatrix}\mathsf{phon} : \text{``newspaper''} \\ \mathsf{syn} : \begin{bmatrix}\mathsf{cat} : \mathsf{n}\end{bmatrix} \\ \mathsf{sem} : \lambda x[\,|\mathbf{newspaper}_U(x)]\end{bmatrix}_U}{\begin{bmatrix}\mathsf{phon} : \text{``newspaper''} \\ \mathsf{syn} : \begin{bmatrix}\mathsf{cat} : \mathsf{n}\end{bmatrix} \\ \mathsf{sem} : \lambda x[\,|\mathbf{newspaper}_U(x)]\end{bmatrix}_U}\ \mathbf{LEX\text{-}NEWSPAPER}$$

As a result of **LEX-NEWSPAPER**, the utterance is assigned a conventional meaning, thus satisfying the IPIH even though this initial interpretation will have to be completed after encountering disambiguating information or at the end of the sentence as discussed in Section §3.4. Note, however, that in order to explain the Frazier and Rayner data it is crucial that no contextual reasoning takes place after the word *newspaper* is encountered, since such reasoning would result in the identification of the appropriate contextual interpretation for *newspaper*, and this in turn would result in garden paths in examples like (15a-d); i.e., the data by Frazier and Rayner appear to contradict the 'incremental understanding' hypothesis proposed in (Chater *et al.*, 1995).[37]

4.4 Priorities among Defaults

If conflicts among defaults are possible, and a lot of words have more than one meaning or sense, why is it that humans do not find most utterances ambiguous? The reason is that, in the case of utterance interpretation just as for other forms of commonsense reasoning, some defaults are stronger than others, and when producing utterances cooperative speakers try to take this into account, thus reducing ambiguity.[38]

Preferences in Commonsense Reasoning

In the case of commonsense reasoning, it has long been known that there are preferences among default inference rules: for example, that more SPE-

[37]Other accounts of lexical access based on underspecified interpretations include the proposal adopted in PHLIQA1 (Bronneberg *et al.*, 1979), Hobbs' proposal (Hobbs and Martin, 1987), and Hirst's 'Polaroid-words' technique Hirst (1987); all of these accounts can be viewed as using p-underspecified representations for both homonyms and polysemous words. In PHLIQA1 and in Hobbs's theory of lexical disambiguation, an ambiguous predicate is assigned as the initial interpretation of both homonyms and polysemous words; Hirst's account is more complex in that it includes an element of 'lazy evaluation', but again, only a single hypothesis is generated, rather than generating multiple ones and then filtering them right away as proposed here. More recently, Copestake and Briscoe (1995) differentiate between homonyms and polysemous words, and propose an account of how polysemous words are disambiguated. Copestake and Briscoe draw a distinction between two types of polysemous words: the cases of CONSTRUCTIONAL POLYSEMY, such as *fast* (that has different interpretations in *fast car* and *fast typist*); and the cases of SENSE EXTENSION–e.g., words like *apple* that have both a mass and a count reading. Two rather different mechanisms are proposed to explain how the senses of these two types of polysemous words are identified. In the case of constructional polysemy, it is proposed that the initial interpretation is p-underspecified; the resolution process they propose is also compatible with what has been proposed here, at least as far as the rather abstract properties of the disambiguation process we are considering here are concerned (a more complex theory of the lexicon is assumed, as well as a different theory of defeasible reasoning). In the cases of sense extension, Copestake and Briscoe propose that lexical rules are used to generate what essentially amounts to new lexical entries for these words; from the point of view adopted here the process by which these cases of polysemy is resolved has therefore similarities to the cases of homonymy we have been considering (except that the separate lexical entries are related in a systematic fashion); and indeed, Copestake and Briscoe claim that these words, as well, can be used for zeugmatic effects as in (9a). We don't know if this similarity has been tested in psychological experiments.

[38]This is not to say that speakers are always cooperative (Davies, 1998).

CIFIC default inference rules tend to be preferred to less specific ones. A simple illustration of the role of specificity is the following example, from Reiter and Criscuolo (1981). Suppose Peter is a student, and all students are adults—i.e., suppose the set of facts of our default theory is:

$$\mathrm{W} = \{\ \mathbf{student}(\mathit{peter}),\ \forall x\ \mathbf{student}(x) \rightarrow \mathbf{adult}(x)\ \}$$

and suppose we have two defaults: that adults are usually employed, and that students are usually unemployed:

$$\mathrm{D} = \left\{ \begin{array}{l} \dfrac{\mathbf{adult}(X)\ :\ \mathbf{employed}(X)}{\mathbf{employed}(X)}\mathbf{ADULT\text{-}EMPLOYED} \\ \dfrac{\mathbf{student}(X)\ :\ \neg\mathbf{employed}(X)}{\neg\mathbf{employed}(X)}\mathbf{STUDENT\text{-}UNEMPLOYED} \end{array} \right\}$$

Intuitively, we would want to conclude that Peter is unemployed, because the fact that he is a student is more 'specific' than the fact that he is an adult.

However, the default inference rules above will give two extensions for the student case, just as in the Nixon case. Various theories have been proposed to explain how such preferences come about. Reiter and Criscuolo suggested a formalization that results in only one extension and is based on the use of so-called SEMI-NORMAL defaults such as **ADULT-EMPLOYED** in (40). These defaults solve the problem without modifying Default Logic, but have a few unpleasant properties, the main of which is that there is no guarantee that a theory including such defaults will have an extension.

(40) $$\frac{\mathbf{adult}(X)\ :\ \mathbf{employed}(X) \wedge \neg\mathbf{student}(X)}{\mathbf{employed}(X)}\mathbf{ADULT\text{-}EMPLOYED}$$

Priorities among defaults, especially those based on frequency, are expressed very naturally in a probabilistic or statistical inference formalism such as the one proposed in Bacchus (1990). As already said, however, there isn't yet a theory unifying such formalisms with the logics used in formal semantics, such as Compositional DRT.[39] We adopted therefore a way of introducing priorities that is simpler and closer to Default Logic, but at the same time sufficient for our needs: Prioritized Default Logic (PDL) (Brewka,

[39] A very popular way to formulate preferences, used especially in abductive approaches, involves assigning numerical 'weights' to assumptions (Hobbs *et al.*, 1993). These methods work well in practice, but again, we are not aware of a clean way to add such rules to the kind of semantic formalisms we are assuming here.

1991). The central idea of PDL is the hypothesis that the set D of defaults is partitioned in 'layers' of defaults of decreasing priority D_1, D_2, etc. A new definition of extension is given, so that the defaults in the layers with higher priority are applied first, thus possibly blocking the application of conflicting defaults of lower priority:

- Let D_i, $i = 1, \ldots, n$ be sets of defaults, and W a set of formulae. E is a PDL-extension of T = (D_1,…,D_n,W) iff there exist sets of formulae E_1,…,E_n such that

  E_1 is an extension of (D_1,W)
  E_2 is an extension of (D_2,E_1)
  ...
  E = E_n is an extension of (D_n,E_{n-1})

With PDL, we can explain the precedence of **STUDENT-UNEMPLOYED** over **ADULT-EMPLOYED** by assuming that defaults are divided in two sets D_1 and D_2 and that **STUDENT-UNEMPLOYED** is in the set of higher-priority defaults, D_1, whereas **ADULT-EMPLOYED** is in the set of defaults of lower priority, D_2.

Preferences in Interpretation

The existence of preferences for certain interpretations in all types of interpretive processes is well-documented. In the case of lexical access, at least two factors have been shown to play an important role (Simpson, 1994). The first factor is FREQUENCY: e.g., Forster and Chambers (1973); Luce (1986) found that the lexical decision task is faster with high-frequency words such as *door* than with low-frequency ones such as *cask*. Equally well-known are so-called SEMANTIC PRIMING effects (Meyer and Schvaneveldt, 1971), i.e., the fact that subjects are faster at the lexical decision task for a pair of words if the words are related (as in *bread* and *butter*) than if they aren't (as in *bread* and *pen*).

The existence of preferences for certain interpretations has also been demonstrated in the case of parsing, although in this case there is more contrast between those who attribute the preferences to a few structurally-based principles (Frazier, 1987) and those who instead claim them to be statistical in nature–e.g., based on the frequency of certain constructions (Hindle and Rooth, 1993; Trueswell *et al.*, 1993; MacDonald *et al.*, 1994). The factors that have been claimed to affect the interpretation of anaphoric expressions include grammatical structure, gender, thematic roles, and topichood, among others (Poesio and Stevenson, To appear).

Frequency Effects

In the rest of the paper, we will show that logics such as PDL can be used to formalize preferences in utterance processing, as well. As an example of how PDL can be used to formalize priorities, let us consider frequency effects. These are rather easy to formalize in PDL: we just have to assume that the default inference rules encoding the most frequent interpretation of certain lexical items occur in sets of defaults with higher priority. For example, if the 'croak like a frog' sense of *croaked* were more frequent than the 'dying' sense, the default inference rule $\mathbf{LEX\text{-}CROAK}_f$ would be in a higher priority class than the default inference rule $\mathbf{LEX\text{-}CROAK}_d$. As a result, $\mathbf{LEX\text{-}CROAK}_f$ would be activated first; and $\mathbf{LEX\text{-}CROAK}_d$ could not apply, because it would result in a contradictory assignment of a value to the function $\rightsquigarrow$. Frequency effects among words (e.g., *door* vs *cask*) could also be explained by assuming that the default inference rules encoding the lexical entries of high-frequency words such as *door* are in a class of higher priority than the lexical entries for low-frequency words such as *cask*.

Semantic Priming

Semantic priming effects can also be explained within PDL by hypothesizing that in addition to the lexical entries $\mathbf{LEX\text{-}CROAK}_d$ and $\mathbf{LEX\text{-}CROAK}_f$ discussed above, whose activation depends only on the occurrence of an utterance of a word, the lexicon also includes default inference rules that depend on contextual factors such the previous utterance of a word whose meaning 'primes' one particular sense, and that these rules have stronger priority than the rules discussed above. Such rules would take the form of $\mathbf{LEX\text{-}CROAK}_f$**-PRIMED** below, where we have used again an utterance-based notation:

$$\dfrac{\begin{array}{l} \mathit{U1}{:}\mathbf{utter}(A,\text{“}\mathit{croaked}\text{”}) \wedge \\ \mathit{U2} \rightsquigarrow \mathbf{P} \wedge \\ \mathbf{primes}(\mathbf{P},\mathbf{croak}_f) \end{array} : \begin{array}{l} \mathbf{syn}(\mathit{U1})\ \mathbf{is} \begin{bmatrix}\mathsf{cat : iv}\end{bmatrix} \wedge \\ \mathit{U1} \rightsquigarrow \mathbf{croak}_f \end{array}}{\begin{array}{l} \mathbf{syn}(\mathit{U1})\ \mathbf{is} \begin{bmatrix}\mathsf{cat : iv}\end{bmatrix} \wedge \\ \mathit{U1} \rightsquigarrow \mathbf{croak}_f \end{array}}\ \mathbf{LEX\text{-}CROAK}_f\mathbf{\text{-}PRIMED}$$

It is generally assumed that lexical priming has to do with closeness in a semantic network (Hirst, 1987); the relation **primes** could then be defined as a disjunction of relations expressing closeness—e.g., subsumption, selectional restrictions, etc.:

$$\forall\ \mathbf{p}\ \mathbf{q}\ \mathbf{primes}(\mathbf{p},\mathbf{q}) \equiv (\forall\ x\ \mathbf{p}(x) \rightarrow \mathbf{q}(x)) \vee (\forall\ x\ y\ \mathbf{q}(x,y) \rightarrow \mathbf{p}(x)) \vee \ldots$$

According to this definition, $\mathbf{croak}_f$ would be primed either by a predicate that subsumes it, or by a predicate that specifies selectional restrictions on its first argument, such as **frog**. **LEX-CROAK**$_f$**-PRIMED** would then be activated in any context in which such a predicate has been uttered.

The formulation just given is only intended as an illustration, not as an exact formalization of lexical priming and / or the related concept of SPREADING ACTIVATION (Hirst, 1987). A formalization closer to current theories of lexical priming would involve rules that make 'concepts' such as the predicate $\mathbf{croak}_f$ 'active' to a certain degree after the activation of the predicate **frog**; in turn, the choice of this particular interpretation of the word *croaked* would be the result of it being more active than alternative interpretations.[40] Nevertheless, the formalization just given does capture one of the main characteristics of priming, namely, the need to make a distinction between concepts that have been mentioned and concepts that haven't. This distinction can be easily made here by requiring, as in **LEX-CROAK**$_f$**-PRIMED**, that the priming concept be the conventional meaning of an utterance in the present context; but it would be difficult or impossible to capture in a theory of utterance interpretation that did not assume that information about what has actually been uttered is available when disambiguation is performed.

In addition, rules such as **LEX-CROAK**$_f$**-PRIMED** can also be used to account for 'parallel lexical disambiguation' effects without introducing additional indexing mechanisms. As noted, e.g., in (van Deemter, 1996; Reyle, 1996), there is a strong tendency to interpret two occurrences of the same word in the same text in the same way. These authors propose constraints on the interpretation of expressions in their logics to achieve this; but the

[40] A theory of lexical access that reflects more closely this property of priming, while preserving some of the desiderata for a theory of disambiguation that we have mentioned in Section §2–the possibility of activating multiple hypotheses, while allowing for incremental processing–has been proposed by Schütze (1997). Schütze also wants to allow for some of the data about disambiguation discussed in Section §2, although he does not seem to make some of the distinctions (e.g., between homonyms and polysemous words). More work is needed to integrate a proposal like his within a more general theory of semantic interpretation like the one proposed here.

observation could equally well be explained as a priming effect, and the result they intend to achieve by means of constraints on interpretation follows already from the hypothesis just discussed about the relative priority among defaults.[41]

Parallelism, Pruning, and Perceived Ambiguity

We can now return to the problem raised in §4.3: how to formalize pruning, which is a crucial ingredient in a parallel theory of interpretation. The answer is now 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.

At the end of each round of hypothesis generation, the language processor may therefore find itself in one of two situations: either there is only one remaining extension, in which case the language processor 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, a p-underspecified interpretation, or an h-underspecified interpretation. The second possibility is that more than one extension remains, because more than one conflicting default inference rules with the same priority could be activated.

However, we also saw in Section §2 that 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 only one is maintained. In the cases of perceived ambiguity discussed in Section §2.5, however, it appears that the conflicting extensions are kept 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 Hopkins' poem. We hypothesize that what happens after an initial pruning according to PDL rules is as follows. If the competing hypothesis are sufficiently strong, they are both 'perceived': this is what happens in Hopkins'

[41] The account we have given can also explain the results by Seidenberg *et al.* (1982); Duffy *et al.* (1988) (reported by (MacDonald *et al.*, 1994)) that whereas equibiased words have long fixation times, non-equibiased words or words in context have shorter fixation times. The idea would be that high priority rules are activated first, whereas low-priority rules are activated later. This may also account for some timing effects (cfr. (MacDonald *et al.*, 1994, page 128)).

poem. Otherwise, one or the other hypothesis may be chosen pretty much randomly: this is what seems to happen (see, e.g., Poesio (1994)) when two weak defaults conflict in the cases of scope assignment studied by Kurtzman and MacDonald (1993), and we hypothesize that the same happens in the Corbett and Chang cases. We can characterize the first situation as one in which two defaults with high priority interact; the second one as the case in which the conflicting defaults have low priority.

If this hypothesis is correct, we may need to revise point (b) of Frazier and Rayner's IPIH–i.e., the claim that the processor cannot delay semantic commitment if this results in maintaining multiple incompatible values for a word, phrase or relation. More psychological evidence is needed, however–in particular, it would be interesting to study the cases in which more than one interpretation seem to be perceived.

One point worth mentioning about the way pruning is modelled in the theory we are proposing is that given the way extensions are defined in PDL (by iterative application of rules of decreasing priority), we appear to predict that in a situation in which more than one default may apply, only the hypotheses produced by the defaults with higher priority will actually be generated: e.g., if one of the lexical entries is more frequent, or is semantically primed, only that hypothesis will be generated. Real parallelism will only obtain in non-primed contexts or in cases when all senses are equally frequent. This would seem consistent with the picture of lexical disambiguation arising from the most recent results (Simpson, 1994); but caution is needed in making such predictions, because the extension generation procedure in PDL shouldn't be taken as a real model of how extensions are generated, but only of the final outcome of this process (i.e., which extensions are going to survive in the end). So it would not be a problem if in the end we were to discover that in fact all extensions are computed, if only briefly, and the priorities only apply later.

Simpson (1994) also points out that there is evidence that the set of applicable defaults, as well, may change as the input gets processed, so that whereas utterance u is interpreted according to the set of default inference rules D, a subsequent utterance u' may be processed according to a different set of defaults D' (Paul *et al.*, 1992). Something like this seems to happen in anaphoric interpretation as well, as we will see in a moment.

5 Processing Pronouns

Pronoun resolution is a well-studied aspect of semantic processing, but the respective roles of incrementality and underspecification for this interpretive process have been discussed much less than in the case of lexical access. In this section we use the theory of semantic processing developed in the previous two sections to account for the evidence about pronouns discussed in Section §2.

The theory we propose has two main ingredients: the claim that the lexical entries of pronouns are semantically h-underspecified, and a formalization of the theory of salience currently most popular, CENTERING THEORY (Grosz *et al.*, 1995), in terms of PDL. We discuss each of these in turn.

5.1 Lexicon and The Conventional Meaning of Pronouns

Extracting a theory of interpretation out of the indications of mainstream semantic theory is even more problematic in the case of pronouns and other 'contextually dependent' expressions than it is in the case of other lexical items, because in this case even the proposals about the conventional meaning of the expression coming from formal semantics cannot be adopted as such. The problem is that with the ordinary tools of semantic theory it is difficult to specify a compositional meaning for these expressions without building a model of context into interpretation; what we get, then, are proposals like Montague's, who derived all possible interpretations of bound pronouns by assigning them one lexical interpretation for each variable x_i in the language (see, e.g., Dowty *et al.* (1981) for discussion), thus making them infinitely ambiguous.[42] Clearly, while we can expect to have a distinct default inference rule for each sense of a lexically ambiguous word (and indeed, psychological research suggests that they may all be activated in non-primed contexts, if they are all equally frequent), generating all of the infinitely many possible interpretations of a pronoun in this way isn't possible.

A way to address the problem is to regard anaphoric expressions as unbound variables: for example, *He left* would translate into **left**(X), where X is not bound (Heim, 1982). This treatment of pronouns neatly captures the intuition that anaphoric expressions are 'context-dependent' by making

[42]In fact, even more interpretations would be needed to allow pronouns to have a referential interpretation in addition to their bound interpretation (Partee, 1972), not to mention the so-called 'e-type' or 'descriptive' interpretation (Evans, 1980; Neale, 1990).

them dependent on the particular assignment with respect to which they are evaluated. However, such translations make the conventional meaning of an utterance containing a pronoun either uninterpretable or logically equivalent to a (possibly infinite) set of propositions, one for each value that may be assigned to that pronoun; either way, no inferences can be made on the basis of that interpretation, so they can only serve as the basis for extra-logical accounts of pronoun resolution (such as, e.g., (Alshawi, 1990)).

None of these objections to the proposals coming from mainstream semantic theory that concern the lexical semantics of anaphoric expressions and do not involve underspecified interpretations is really insurmountable. However, the possibility that lexical entries may be underspecified opens the door to ways of dividing the labour between lexicon and context that we find more intuitive, and do not force us to stipulate implausible ambiguities or to introduce uninterpretable objects in our logic.

More specifically, on the basis of the discussion in Section §3.4 we can entertain two hypotheses concerning the semantics of pronominal expressions: that the conventional meaning of pronouns is p-underspecified, or that it is h-underspecified. In fact, the hypothesis that pronouns have a p-underspecified conventional meaning has already been made in formal semantics, without explicitly mentioning underspecification. One way to implement this idea within a DRT framework is to claim that the pronoun *he* introduces a new discourse entity x_n, known to be male, as in (41). This makes the pronoun semantically equivalent to an existential quantifier; establishing a co-specification relation with an antecedent would be a case of pragmatic strengthening, along the lines suggested in (Kripke, 1977) for definite descriptions. This interpretation is p-underspecified in the sense that the utterance containing the pronoun is given a conventional meaning that subsumes all of its possible interpretations.

(41) $$u_{pro}\text{:}\mathbf{utter}(a,\text{“}he\text{”}) \wedge \mathbf{syn}(u_{pro}) = \begin{bmatrix} \mathsf{cat} : \mathsf{pro} \\ \mathsf{agr} : \begin{bmatrix} \mathsf{gen} : \mathsf{mas} \\ \mathsf{num} : \mathsf{sing} \\ \mathsf{pers} : 3 \end{bmatrix} \end{bmatrix} \wedge$$

$$u_{pro} \leadsto \lambda P[x_n|\mathbf{male}(x_n)]; P(x_n)$$

A variant of this proposal is the claim that pronouns are disguised definite descriptions referring to the 'most salient' entity (Neale, 1990; Hwang and Schubert, 1993). This approach is not easy to implement within a traditional DRT framework, in which indefinites and definites are not treated as quantifiers, but can be formulated within other dynamic theories in which definite descriptions are treated as quantifiers, such as DPL (Groenendijk and Stokhof, 1991), and could be formalized in Compositional DRT as well, using the type theory in terms of which the DRS constructs are defined.[43]

A more radical alternative is to propose that the lexical interpretation of pronouns is h-underspecified with respect to meaning, rather than p-underspecified. According to this theory, the lexical interpretation of *he* does not specify a conventional meaning at all. We show in (42a) the proposed lexical entry for *he*, in (42b) its translation in utterance notation:

(42) a.
$$\begin{bmatrix} \mathsf{phon}: \text{“he”} \\ \mathsf{syn}: \begin{bmatrix} \mathsf{cat}: \mathsf{pro} \\ \mathsf{agr}: \begin{bmatrix} \mathsf{gen}: \mathsf{mas} \\ \mathsf{num}: \mathsf{sing} \\ \mathsf{pers}: 3 \end{bmatrix} \end{bmatrix} \end{bmatrix}_{u_{pro}}$$

b.
$$[u_{pro} \mid u_{pro}\text{:}\mathbf{utter}(a,\text{“}he\text{”}), \mathbf{syn}(u_{pro})\ \mathbf{is} \begin{bmatrix} \mathsf{cat}: \mathsf{pro} \\ \mathsf{agr}: \begin{bmatrix} \mathsf{gen}: \mathsf{mas} \\ \mathsf{num}: \mathsf{sing} \\ \mathsf{pers}: 3 \end{bmatrix} \end{bmatrix}]$$

Under this hypothesis, determining the conventional meaning of u_{pro} would be entirely left to contextual interpretation: thus, for example, if u_{pro} followed the utterance of *Kermit croaked* discussed in the previous sections, in which the utterance of *Kermit* gets the interpretation in (25), and if *he* was taken to be co-indexed with the utterance of *Kermit*, we could conclude that u_{pro} has the same conventional meaning as u_{pn} in (25):

(43) $u_{pro} \leadsto x_n$

Both (41) and (42) eliminate the need for stipulating multiple ambiguities in the lexicon; the ambiguity is located entirely in the resolution process. But we believe that both psychological and linguistic evidence support the hypothesis that lexical entries of pronouns are h-underspecified with respect

[43]Yet another alternative is the proposal by Kamp and Reyle (1993) that pronouns introduce discourse entities together with semantically uninterpreted conditions of the form $x = ?$ indicating that the discourse entity co-specifies with another discourse entity. This alternative only really works within a ‘representational’ framework, i.e., one in which it is assumed that the conventional meaning assigned to an utterance is actually a representation – an object whose semantics is related to the notion of truth only indirectly, and in which it is therefore allowed to have objects which may not be interpretable in such a way.

to the meaning. We note, first of all, that the proposal in (41) is inconsistent with the evidence about pronoun interpretation discussed in Section §2. If Frazier and Rayner's Immediate Partial Interpretation Hypothesis were correct, and if the lexical entry of pronouns were p-underspecified, we would expect pronouns to behave like polysemous words, i.e., like *newspaper*; we would not expect multiple interpretations of pronouns (and garden paths) to be available in a context in which more than one antecedent satisfies gender constraints. But the data from Corbett and Chang (1983); Gernsbacher (1989); Stevenson and Vitkovitch (1986) suggest that these multiple interpretations are in fact generated. This can mean one of two things: either that the lexical meaning of a pronoun is fully specified (i.e., the lexicon encodes all of the possible interpretations of the pronoun, as in Montague's proposal), which seems very unlikely; or that its conventional meaning is h-underspecified, and the contextual rules that determine its meaning are activated on-line as soon as the pronoun is encountered.

The hypothesis that the lexical meaning of pronouns is p-underspecified is also problematic from a linguistic point of view. The first problem is making sure that this p-underspecified interpretation is flexible enough to cover all of the possible interpretation. Besides ensuring that both the referential and the bound anaphora interpretation are covered, one has to worry about sentences containing 'paycheck pronouns' such as (44):

(44) *The man who gave his paycheck to his wife is wiser than the man who gave <u>it</u> to his mistress*

which suggest that *it* can have a 'type' interpretation as well as a 'token' interpretation; this interpretation could not be covered by a translation along the lines of (41) – say, $\lambda P[x_n|\mathbf{thing}(x_n)]; P(x_n)$ – because the antecedent is actually a type. The data about anaphora to abstract objects (Asher, 1993) also indicate that the pronoun *it* can refer, in additions to entities in the universe, to objects as different as propositions, events, or facts. This suggests that in the case of neuter pronouns at least, it would be difficult to find a conventional meaning general enough to cover all possible cases.[44]

[44] A technical tool recently introduced in formal semantics, type shifting Partee and Rooth (1983) may offer a solution here. The idea is that certain lexical items may have more than one conventional meaning, which are, however, related by systematic operations–in other words, one conventional meaning can be derived from another by applying an operation p to it. In the case of pronouns just discussed, one might argue that they do have only one 'basic' interpretation, and the other ones are obtained if needed by applying type-shifting operations.

There are then more specific problems depending on which version of the p-underspecified hypothesis we adopt. The translation in (41) makes pronouns semantically equivalent to indefinite NPs; we lose thus the opportunity to give a semantic account for definiteness effects (Reuland and ter Meulen, 1987). The 'pronouns-as-descriptions-referring-to-the-most-salient-entity' hypothesis suffers instead from the problem that it's not at all clear that 'the most salient entity' always refers uniquely, even when taking gender into account (Poesio and Stevenson, To appear). Consider the following text (from *The Guardian*, August 15, 1998, Weekend section, p.22): what is the 'most salient (masculine) entity' when the underlined pronoun is to be interpreted?

(45) *He recalls his mother trying to shield him from his father's excesses. "Your father doesn't mean it," she would console him. "He loves you, he is a good man." And for years he thought she was making excuses. "But she wasn't. He is a good man." Just a product of his time. When Maupin was born, his father was in the thick of battle, the skipper of a minesweeper. He didn't see his son for two years. <u>He</u> learned of his birth from a sailor on another ship, by semaphore. "I got very sentimental about [this] six months ago, and asked him to tell me exactly where he was when he found out. And he rather touchingly sent him maps.*

There are also objections to the h-underspecified translation in (42), of course. Some may object to the idea that the lexicon does not specify any restriction on the meaning of a pronoun. We are not claiming that pronouns can mean anything, however; only that their meaning is specified by contextual rules rather than lexical rules. We will see examples of such rules in a moment, but we can already say that such rules are also going to be default inference rules, just like lexical entries are; so all we are claiming is that two distinct sets of defaults–some operating only on the basis of the fact that a certain utterance occurred, other ones using information about the context of utterance–interact to determine the meaning of pronouns.

The translation in (42) could also be criticized for the opposite reason; i.e., some may wonder why we want to draw a separation between the lexicon and context at all, i.e., and why we need to assume two separate sets of defeasible inference rules–some specifying the 'lexical entry' for a pronoun, other specifying how its interpretation is determined in a context–instead of simply assuming the latter as our lexical entries. The answer to this objection has partly to do with modularity arguments–there is increasing evidence for a lexical module independent from the rest of the processing system and that gets activated very early (Corley, 1998)–and partly with the data reported by Garrod and Sanford (1994): if lexical entries for pronouns always access context, one could not explain why pronouns referring to objects 'in topic' are always interpreted, whereas other and pronouns are not (unless we assume that the lexicon is not always accessed). The account of pronoun resolution we are going to present next does keep these issues into account.[45]

5.2 Pronoun Resolution

The evidence discussed in Section §2 suggests that the way pronouns are resolved is in many respects similar to the way homonyms are resolved. More specifically, the results of Corbett and Chang (1983); Gernsbacher (1989) suggest that the h-underspecified interpretation resulting from lexical access (as in (42)) is, in turn, the starting point for a second phase of

[45] Another issue raised by the analysis of anaphoric expressions just presented is: what is the difference between pronoun *it* and semantically vacuous (pleonastic) *it* as in *it rains*? For, if our claims about the lexical entries of pronouns are correct, in both cases we would have lexical entries that do not specify the conventional meaning of an utterance.

We don't have a complete solution to the problem, but there are at least two possible ways of addressing the issue. It may be the case that expletives are not 'semantically empty' in the same sense that pronouns are, but do have a conventional meaning, although one that does not contribute to the overall meaning of the utterance, which would be entirely specified by the predicate. If we assume that *rains* is a zero-argument predicate so that *it rains* translates as

$$[s|s{:}\mathbf{rain},\ s \subseteq S\ ]$$

then we could assume this to be the conventional meaning of *rains*, whereas the conventional meaning of expletive *it* might be of the form $\lambda\varphi.\varphi$, where φ is the proposition denoted by *rains*.

A second hypothesis is that the meaning functions $\rightsquigarrow$ are only partial. Not all utterances would then have a conventional meaning, but only a subset; existence axioms would specify which type of utterances do have a conventional meaning of a certain type. If this were the case, one could then hypothesize that one such axiom holds for utterances of category `pro`:

AX-PRO-TOTAL $\forall u_\epsilon \mathbf{syn}(u_\epsilon) = \begin{bmatrix}\mathsf{cat} : \mathsf{pro}\end{bmatrix} \rightarrow \exists y_e\, u \rightsquigarrow_e y$

but no such axiom would hold for utterances of category `expl`. This solution has two attractive features. First of all, it makes sense that not all utterances have conventional meanings of all types. Secondly, one could provide an account of what happens in a case of 'ambiguity' in which it is very difficult to decide what disambiguation is preferred, if any–namely, the choice between interpreting *it* as an expletive or as a reference to events in examples such as *John was late. It seemed strange to everybody*. The hypothesis would be that in these cases the reader cannot choose between two very similar lexical entries, both equally likely.

interpretation. During this phase of contextual processing, multiple default inference rules suggesting alternative interpretations for the pronoun are activated, when no other factor indicates a priority of certain interpretations over others; if some such factor does exist, higher-priority default inference rules suggest the preferred interpretation in context.

The differences between lexical access and pronoun resolution are, first of all, that the default inference rules involved in pronoun resolution do not encode grammatical knowledge, but resolutions available in context; and second, that the priority ranking among interpretations is determined by the salience of discourse entities, rather than by the frequency of certain interpretations or by priming.

Basic rules for pronoun resolution

We begin by discussing the low-priority default inference rules that generate the interpretation of a pronoun when no discourse entity is most salient. In order to do this, we need to specify more precisely what we expect the final interpretation of pronouns to be after resolution, and how it is obtained. In keeping with the DRT framework, we assume that pronouns are eventually resolved to discourse entities, as in (43).[46] As for the way these values are established, we have to take into account the fact that pronoun interpretation takes place incrementally, i.e., before a complete DRS for the present utterance has been constructed. We propose that pronoun resolution works at the level of utterance events: the contextual resolution rules find appropriate antecedents for an event of uttering a pronoun NP among the events of uttering NPs that are currently most salient. Once an antecedent utterance event has been identified, its associated discourse entity is suggested as a potential conventional meaning for the pronoun.

In order to formalize this proposal, we need first of all to clarify this notion of 'associated discourse entity'. The conventional meanings of NPs discussed so far do not give us discourse entities (see, e.g., the lexical entry for *Kermit* in (1)). In order to ensure that we can always identify the discourse entity introduced by an NP we have to assume that utterances of NPs have an additional feature, which we will call **dref**, specifying the discourse entities introduced into discourse by these utterances. (As it turns out, this is also needed for the treatment of scope in Section §6.) This as-

[46]In a DRT framework, both referential pronouns and anaphoric pronouns get translated into discourse entities; the referential properties of, say, deictic pronouns are mediated by discourse entities and captured by assuming a separate 'anchoring' mechanism for them (Kamp, 1990).

sumption leads to the following slightly revised lexical entries for proper names and indefinites:[47]

(46) $$Kermit_n \rightsquigarrow \begin{bmatrix} \mathsf{phon} : \text{“kermit”} \\ \mathsf{syn} : \begin{bmatrix} \mathsf{cat} : \mathsf{pn} \end{bmatrix} \\ \mathsf{sem} : \lambda \mathsf{P}[x_n | x_n \ \mathbf{is}\ k]; \mathsf{P}(x_n) \\ \mathsf{dref} : x_n \end{bmatrix}_{u_n}$$

(47) $$a_m \rightsquigarrow \begin{bmatrix} \mathsf{phon} : \text{“a”} \\ \mathsf{syn} : \begin{bmatrix} \mathsf{cat} : \mathsf{det} \end{bmatrix} \\ \mathsf{sem} : \lambda \mathsf{P} \lambda \mathsf{Q}[x_m |]; \mathsf{P}(x_m); \mathsf{Q}(x_m) \\ \mathsf{dref} : x_m \end{bmatrix}_{u_m}$$

Secondly, we need to specify which events of uttering NPs may cause the language processor to generate hypotheses about pronoun antecedents. For this, we resort to terminology from Centering theory (Grosz *et al.*, 1995). According to this theory, language users maintain a LOCAL FOCUS of attention which serves as the preferential source for interpreting pronouns, as discussed below. One of the constituents of the local focus which are available when processing C-UTTERANCE u_{n+1},[48] are the discourse entities introduced by c-utterance u_n; these are called FORWARD LOOKING CENTERS in Centering theory, and denoted by $C_f(u_n)$. For example, processing c-utterance u_1 of the string *Kermit likes Miss Piggy* involves creating a list of forward looking centers $C_f(u_1)$ = [k,m_p], which serves as the local focus for the next c-utterance.

In order to make these ideas truly formal, one would have to specify what counts as a c-utterance, and what is the previous c-utterance of a given c-utterance; both of these questions are, however, significant research problems on their own (Kameyama, 1998), that we are not going to solve here. All we need for our purposes is a way of indicating the function mapping c-utterance u into its previous c-utterance u'; we will use the notation **prev-utt**$(u) = u'$. We also use the following notation for talking about forward looking centers. The relation $\mathbf{C}_f(u_n,x)$ holds if x is one of the

[47]Note also that the value of **dref** is not the value stored in the discourse entity (i.e., the value of discourse entity x_n at assignment i) but the discourse entity itself; in Compositional DRT, discourse entities are constants, hence are always 'accessible'.

[48]Centering theory relies on a technical notion of 'utterance' that has nothing to do with the notion of 'utterance' that we have been using; we will therefore use the term C-UTTERANCE to refer to utterances in the sense of centering theory.

forward-looking centers of c-utterance u_n. The relation **cf-utt**(u,u') holds if u' is an utterance of one of the forward-looking centers of u – i.e., if u is a c-utterance and u' is the utterance of an NP as part of u:[49]

$$\forall u, u' \ \textbf{cf-utt}(u, u') \equiv (\textbf{c-utterance}(u) \wedge \textbf{syn}(u') = \begin{bmatrix}\mathsf{cat} : \mathsf{np}\end{bmatrix} \wedge u' \lhd u)$$

We can now specify the basic PDL rule for pronoun resolution, **PRO-MATCH**. This rule states that the language processor may hypothesize that the conventional meaning of the utterance of a pronoun is the same as the conventional meaning of one of the utterances introducing the forward looking centers of the previous c-utterance, provided that their agreement features match.

$$\frac{\begin{array}{ll} [\mid \textbf{syn}(U_{pro}) \textbf{ is } \begin{bmatrix}\mathsf{cat} : \mathsf{pro}\end{bmatrix}, & \\ \textbf{cf-utt}(U_{n+1},U_{pro}),\ \textbf{cf-utt}(U_n,U_{np}), & : \ [\mid U_{pro} \leadsto x] \\ \textbf{prev-utt}(U_{n+1}) \textbf{ is } U_n, & \\ \textbf{agr-match}(U_{np},U_{pro}),\ U_{np} \leadsto x\ ] & \end{array}}{[\mid U_{pro} \leadsto x]}\ \textbf{PRO-MATCH}$$

PRO-MATCH predicts that when there is a single U_{np} matching the agreement features of U_{pro} in the previous c-utterance, a single hypothesis will be generated; whereas more than one hypothesis will be obtained when there is more than one antecedent. The rule also shows how syntactic factors may play a role in pronoun interpretation; we note that this is hard to do in theories that, unlike ours, do not explain how syntactic information becomes available during interpretation.[50]

Centering Theory As A Source of Preferences in Pronoun Interpretation

Of all the factors that play a role in determining a preference for one interpretation over the other ones in the case of pronoun resolution, we concentrate here on those discussed by (Garrod and Sanford, 1994), and in particular, the relative salience of discourse entities. Several theories of salience

[49]We are making the simplifying assumption that every NP in a c-utterance introduces a discourse entity.

[50]Corbett and Chang (1983) actually observed that hypotheses about pronoun interpretation may not be filtered out on the basis of agreement only - i.e., that hypotheses concerning antecedents not matching in gender were also generated. It's not clear what the role of number may be, if any.

have been proposed (Poesio and Stevenson, To appear); we adopt here a version of Centering theory.[51]

We already introduced some of the terminology used in Centering theory, and in particular the notions of forward-looking centers and c-utterance. The other crucial notion of the theory is the BACKWARD LOOKING CENTER (C_b), another component of the 'local' focus of attention. The C_b of c-utterance u_{n+1} is defined as the highest-ranked C_f of **prev-utt**(u_{n+1}) (= u_n) that is realized in u_{n+1}. Precise definitions of the notions of 'ranking' and 'realized' are still open questions in Centering theory, just like the notion of c-utterance introduced above;[52] we will ignore these problems, as well, and simply assume that there is a function **rank**(u,x) determining the rank of discourse entity x in utterance u. We also need a second function, **highest-ranked**(u), defined as follows:[53]

$$\forall u_n,\ x\ (\textbf{highest-ranked}(u_n) = x) \equiv$$
$$(C_f(u_n,x) \wedge (\forall x'\ (x \neq x' \wedge C_f(u_n, x')) \rightarrow \textbf{rank}(u_n, x') < \textbf{rank}(u_n, x))$$

The function $\mathbf{C}_b(u)$ can also be defined in terms of **rank** as follows:

$$\begin{array}{l}\forall u_n, u_{n+1},\ \textbf{prev-utt}(u_{n+1}) = u_n \rightarrow \\ \quad (\forall\, x\, C_b(u_{n+1}) = x \equiv \\ \quad\quad (C_f(u_{n+1}, x) \wedge C_f(u_n, x) \wedge \\ \quad\quad (\forall x'\ (x \neq x' \wedge C_f(u_{n+1}, x') \wedge C_f(u_n, x')) \rightarrow \\ \quad\quad \textbf{rank}(x', u_n) < \textbf{rank}(x, u_n))))\end{array}$$

The main claim of Centering theory as far as pronominalization is concerned is that if anything is pronominalized in an utterance, the C_b is. The one algorithmic formulation of Centering theory (Brennan *et al.*, 1987) implements this idea by means of a generate-and-test algorithm that generates all possible interpretations of a pronoun and then uses this (and other) claims of Centering theory as filters; but the claim can also be formulated in terms of the theory of semantic interpretation and of prioritized default

[51]We note that the theory of interpretations proposed in Section §3 satisfies the original desiderata for a theory of interpretation supporting Centering theory–see (Grosz *et al.*, 1983).

[52]Most versions of the theory assume that ranking is determined by grammatical factors. This hypothesis is especially explicit in one of the latest versions of the theory, proposed in Gordon *et al.* (1993) on the basis of psychological experiments; in this version, the C_b of an utterance is simply identified with its subject.

[53]Experimental results by, among others, Gordon *et al.* (1993) suggest that ranking may be a partial rather than a total order. We will ignore this problem here.

rules proposed in this paper. The simplest way to do this would be to follow the suggestion in (Gordon, 1993, page 41):

> ...In particular, [Centering theory] states that a syntactically ambiguous pronoun will be interpreted by default as realizing the highest ranked member of the set of forward-looking centers of the previous utterance until such time as that interpretation is overridden by other information.

This suggestion can be implemented by means of the following default:

$$\frac{\begin{array}{l}[|\ \textbf{syn}(U_{pro})\ \textbf{is}\ \begin{bmatrix}\textsf{cat : pro}\end{bmatrix}, \\ \textbf{cf-utt}(U_{n+1},U_{pro}),\ \textbf{cf-utt}(U_n,U_{np}), \\ \textbf{prev-utt}(U_{n+1})\ \textbf{is}\ U_n, \\ \textbf{agr-match}(U_{np},U_{pro}),\ U_{np} \rightsquigarrow x, \\ \textbf{highest-ranked}(U_n)\ \textbf{is}\ x\ ]\end{array} \quad : \ [|U_{pro} \rightsquigarrow x]}{[|U_{pro} \rightsquigarrow x]}\ \textbf{PRO-CB-GORDON}$$

The problem with **PRO-CB-GORDON** is that it predicts that there is going to be a preference for a given discourse entity for every c-utterance in a text except for the very first one, since every c-utterance has a highest-ranked C_f. However, the results of Corbett and Chang (1983); Gernsbacher (1989) suggest that this is not the case; with materials like those they use, at least, all discourse entities agreeing with the pronoun are activated in the second c-utterance (see (12), repeated below for convenience).

(12) *Karen poured a drink for Emily and then she put the bottle down*

We believe that the reason for this is that focusing effects only begin to manifest themselves after the C_b has been established at least once, i.e., starting from the *third* c-utterance, as often suggested in the literature on focusing and centering (Sidner, 1979; Poesio and Stevenson, To appear). The materials used by Corbett and Chang (in (12)) are such that the pronoun occurs in the second c-utterance of the text, which means that no C_b is defined for the previous c-utterance.

We propose therefore that the default proposed by Gordon should be revised as a preference to interpret pronouns as referring to the C_b of the previous c-utterance, rather than simply the most highly-ranked element of the previous c-utterance realized in the current c-utterance, as follows:

> A syntactically ambiguous pronoun will be interpreted by default as realizing the C_b of the previous utterance until such time as that interpretation is overridden by other information.

This default may be encoded in PDL as follows:

$$\frac{\begin{array}{l}[|\ \textbf{syn}(U_{pro})\ \textbf{is}\ \begin{bmatrix}\mathsf{cat : pro}\end{bmatrix}, \\ \textbf{cf-utt}(U_{n+1},U_{pro}),\ \textbf{cf-utt}(U_n,U_{np}), \\ \textbf{prev-utt}(U_{n+1})\ \textbf{is}\ U_n, \\ \textbf{agr-match}(U_{np},U_{pro}),\ U_{np} \rightsquigarrow x, \\ \textbf{CB}(U_n)\ \textbf{is}\ x\ ]\end{array} \quad : \ [|U_{pro} \rightsquigarrow x]}{[|U_{pro} \rightsquigarrow x]}\ \textbf{PRO-CB}$$

This default would then have to be supplemented with (weaker) defaults encoding the effects of parallelism and, in particular, the prominence of subjects to account for the data of (Gordon *et al.*, 1993; Stevenson *et al.*, 1993; Smyth, 1994); other factors have been showed to play a role in pronoun interpretation as well. We will not try to do this here.

Two Examples

To see how these rules work, let us first consider a context similar to those studied by Corbett and Chang, in which an utterance u_1 of *Kermit met Hobbes* is followed by an utterance u_2 of *He croaked*. The relevant facts for centering theory holding after u_1 are shown in (48); u_{k_1} is the utterance of *Kermit* as part of u_1, introducing discourse entity x_{k_1}, whereas u_{h_1} is the utterance of *Hobbes* as part of u_1, introducing discourse entity x_{h_1}.

(48) $\mathbf{C}_f(u_1,x_{k_1}) \wedge \mathbf{C}_f(u_1,x_{h_1})$

The crucial bit is that since u_1 has no previous c-utterance, $\mathbf{C}_b(u_1)$ is not defined. Now, let us consider what happens when the pronoun *He* is uttered at the beginning of u_2. Let u_{pro} be the utterance of *He* in u_2. The result of lexical access for the pronoun is as for the lexical entry in (42b):

(49) $[u_{pro}|\ u_{pro}\textbf{:utter}(a,\text{“}he\text{”}),\ \textbf{syn}(u_{pro})\ \textbf{is}\ \begin{bmatrix}\mathsf{cat : pro} \\ \mathsf{agr :} \begin{bmatrix}\mathsf{gen : mas} \\ \mathsf{num : sing} \\ \mathsf{pers : 3}\end{bmatrix}\end{bmatrix}\ ]$

If we start defeasible reasoning at this point, we get two distinct extensions. One of them is obtained using **PRO-MATCH** with $U_{np} = u_{k_1}$; this extension includes the fact in (50a). The other extension is obtained using **PRO-MATCH** with $U_{np} = u_{h_1}$, and includes the fact in (50b).

(50) a. $u_{pro} \rightsquigarrow x_{k_1}$

b. $u_{pro} \rightsquigarrow u_{h_1}$

Note that at this point **PRO-CB** does not apply, because $\mathbf{C}_b(u_1)$ is not defined. At this point therefore the two extensions are equivalent, and the processor cannot resolve the ambiguity; both hypotheses remain 'alive' until the end of u_2 is reached, just as observed by Corbett and Chang. One of them may then be eliminated on the basis of selectional restrictions.

Let us consider now a case in which the C_b is established before encountering the pronoun. Assume that *Kermit is a frog* (let us call this utterance u_0) had been uttered before *Kermit met Hobbes* (u_1). The facts holding after u_0 are as follows:

(51) $\mathbf{C}_f(u_0, x_{k_1})$

After u_1, the same facts hold as in the previous example ((48)) except that now the C_b of u_1 is defined,

(52) $\mathbf{C}_b(u_1) = x_{k_1}$

since x_{k_1} is the only discourse entity that is a C_f both of u_0 and of u_1. Now, let u_{pro} be, again, the utterance of *He* in u_2. At this point, **PRO-MATCH** could apply again, generating again extensions including the facts in (50). However, the generation of these hypotheses is prevented by the fact that the higher-priority PDL rule **PRO-CB** can apply, producing a single extension containing (50a). Just as in the examples of lexical disambiguation seen in Section §4, higher-priority rules block the application of lower-priority ones.

The Results of Garrod and Sanford

We conclude this section by returning to the result reported by Garrod and Sanford (1994), that whereas definite descriptions and proper names are always interpreted, pronouns only seem to be interpreted when referring to the 'topic', which we will identify with the C_b here.

Within the formalization we have proposed, we can now explain these results as another instance of the phenomenon discussed by Simpson (1994) for lexical access–namely, that not all default inference rules are always active. More specifically, we hypothesize that high-priority rules are always active, whereas low-priority ones are only active when the task is really important and the language processor is 'trying really hard'. The data reported by Garrod and Sanford suggest that of the rules defined above, **PRO-CB** is always active, whereas the lower-priority **PRO-MATCH** is optional. If only high-priority rules are active, the language processor, when encountering (16e) or (16f) can only resolve *she* (using **PRO-CB**), but not *he*, because **PRO-MATCH** doesn't apply. In this latter case, the conventional meaning of the pronoun would seem to remain h-underspecified, and as a result, the conventional meaning of the utterance as a whole would remain h-underspecified, as well. (Or, perhaps, some sort of simplified meaning may be assigned to the utterance in these cases– but it's not clear what such a meaning could be.)

6 Scope

A lot of work on underspecification in semantics has been motivated by the problem of scope assignment. There are several good reasons for this, the most important being that in this case, more than in the cases of lexical disambiguation and pronoun resolution, the combinatorics are horrendous: as shown by the example (7) presented in the Introduction, the number of scopally distinct interpretations may run in the thousands. This strongly suggests that, unlike lexical ambiguities, scopal ambiguities are not resolved by generating all interpretations and choosing among them. A second reason why scope assignment is interesting is that unlike in the case of lexical disambiguation and pronoun resolution, in the case of scope assignment subjects do not always seem to have strong intuitions. Because of this, many researchers hypothesizing suspect that scope assignment is a good example of ambiguity left unresolved, except perhaps in the case of sentences containing only one or two quantifiers, and 'simple' ones (van Benthem, 1982).

While this state of affairs makes a discussion of the role of semantic underspecification in scope assignment unavoidable, the problems in eliciting subjects' judgment mean that, as already said in Section §2, the psychological evidence for scope assignment is nowhere as abundant as in the case of lexical disambiguation and pronoun resolution; in fact, we know of only two published studies, (Ioup, 1975; Kurtzman and MacDonald, 1993). This prevents the development of a complete theory of scope assignment at this stage; for this reason, the hypotheses presented in this section are rather more tentative than those about lexical disambiguation and pronoun resolution.

In this section we will first of all discuss the form of underspecified interpretation that is passed from grammar to contextual processing, then how this underspecification may be resolved.

6.1 Scopal Underspecification

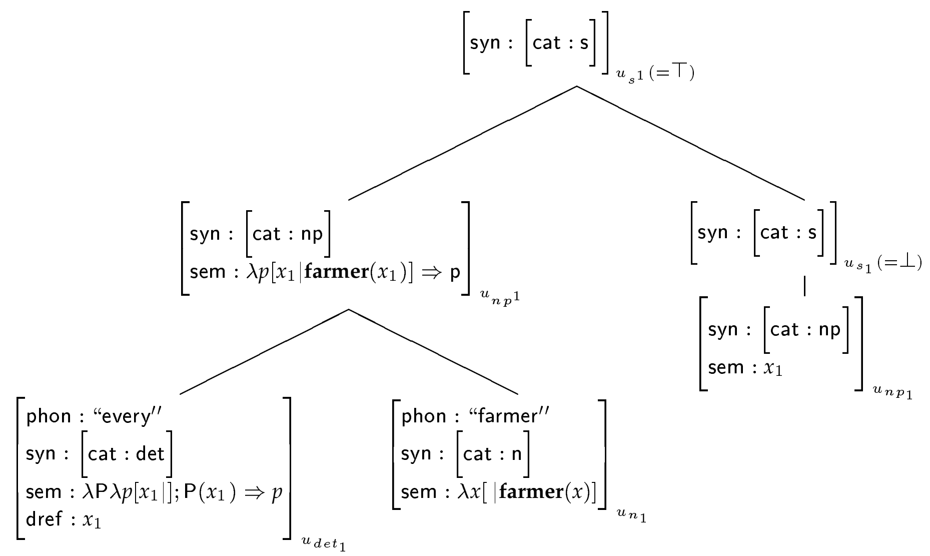
The approach to scope underspecification we adopt here[54] builds on two hypotheses. The first of these is that the structural relations among utterances inferred by the language processing module do not amount to an s-structure, but reflect instead the structure of the derivation of the conventional meaning–i.e., they specify what is called a 'derivation tree' in Montague grammar or 'LF' in generative grammar. (See also the discussion of Nunberg's proposal in Section §4.) For example, arriving at the (completely disambiguated) conventional meaning of *every farmer owns a donkey* in which *every farmer* takes wide scope over *a donkey* involves inferring the structural relations shown in Figure 4.

Figure 4 also illustrates our second important hypothesis, which actually dates back to Robin Cooper's dissertation (Cooper, 1975).[55] This is the idea that a (quantified) NP contributes twice to the derivation of the conventional meaning of an utterance: once by specifying an argument of the verb, and a second time as an operator taking a sentential argument. The NP *every farmer*, for example, contributes both the utterance u_{np_1} and the utterance u_{np^1} to the derivation of the meaning of *every farmer owns a donkey* illustrated in Figure 4. (We use the convention of assigning the same index to these utterances, and indicating the first of them with a subscript index, the second with a superscript one.) The semantic value of utterance u_{np_1} is the discourse entity x_1; u_{np^1} contributes the compositional meaning usually associated with quantified NPs, which in Compositional DRT is rendered as $\lambda p[x_1|\mathbf{farmer}(x_1)] \Rightarrow p$. The second utterance's contribution to the meaning corresponds to what is 'put in store' in Cooper's treatment. Within an LTAGs framework, the contribution of a quantified NP to the derivation of the conventional meaning of an utterance can be formalized as a single tree. More specifically, the LTAG tree introduced by *every farmer* is shown in (53).

[54]Many of the ideas presented here have been inspired by Reinhard Muskens, who has been independently developing a theory of scope underspecification presented in (Muskens, 1995a) as well as in work still unpublished.

[55]See also (Pereira, 1990; Carpenter, 1994).

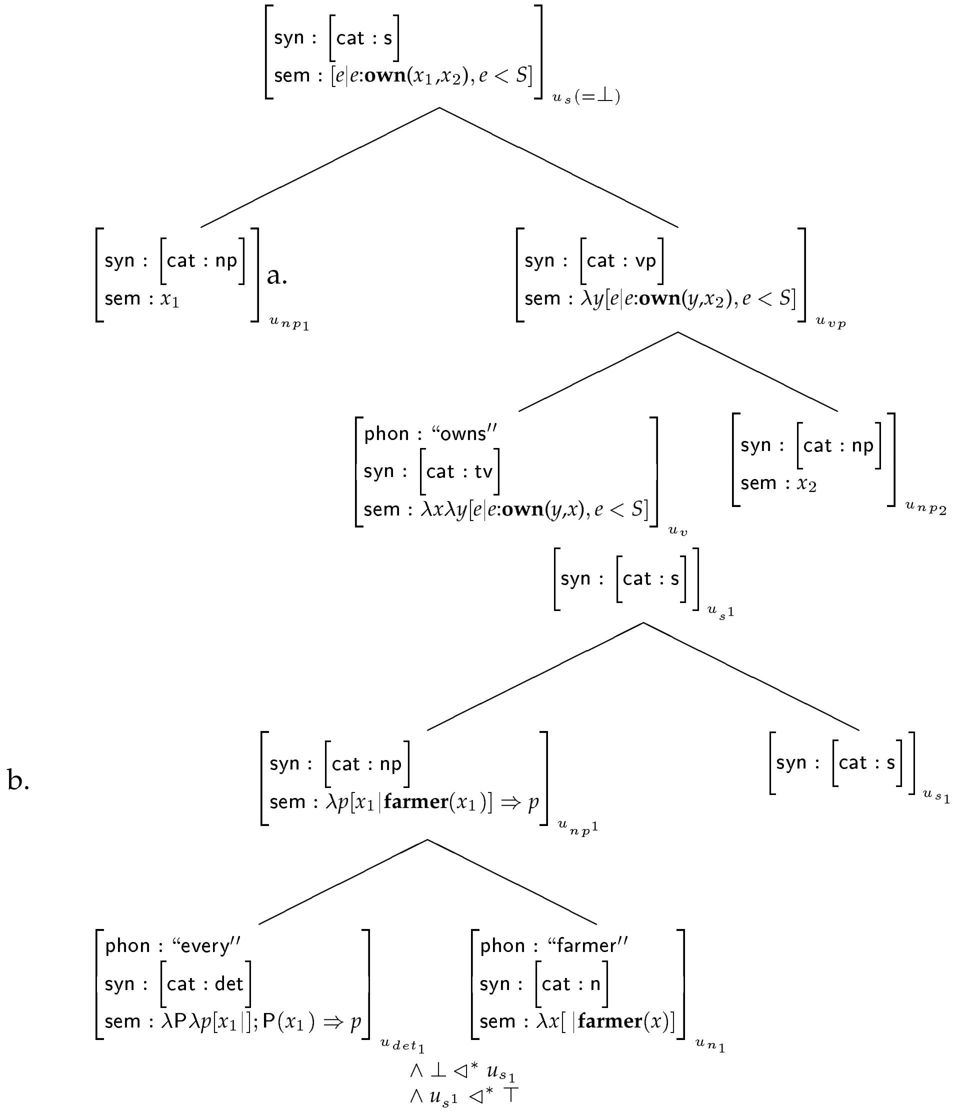
(53)



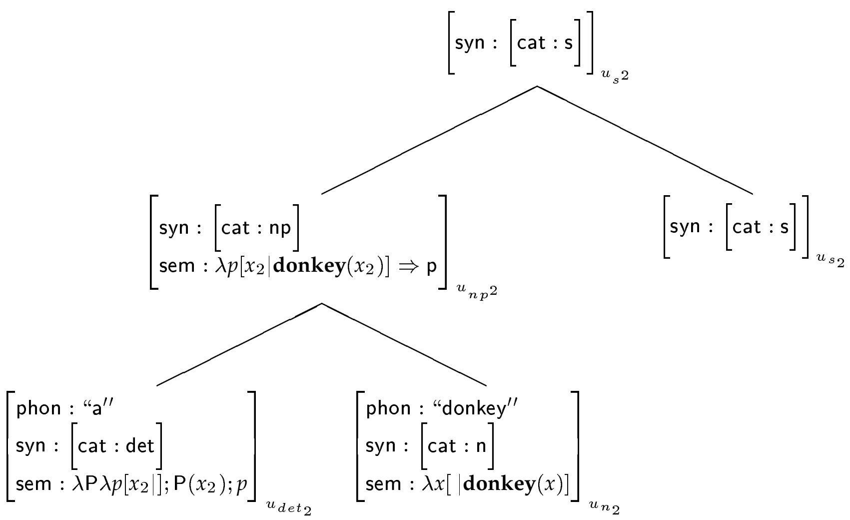
If we accept these two hypotheses, scope underspecification becomes a special case of structural underspecification, now seen more precisely as a partial specification of meaning derivation; the underspecification is due to the fact that of the two utterances that result from perceiving an NP, only the one contributing an argument to the verb (u_{np_1} in the example we have been considering) has a fixed position within the utterance decomposition structure. The position within the derivation tree of the utterance contributing a quantifier (u_{np^1}), on the other hand, is not determined by grammatical knowledge; all that grammar specifies about its role in the derivation is that the scope of the quantifier is obtained by a derivation process that involves at the very least determining the arguments of the matrix verb (the point at which this happens is indicated in Figure 4 as $\perp$), and that in turn the result of applying the quantifier to this scope is part of the decomposition of the utterance as a whole (indicated in Figure 4 as $\top$). The constraints on the derivation of the conventional meaning of the utterance *every farmer owns a donkey* prior to scope assignment are summarized by the partial description in (54): the partial derivation in (54a) specifies how the meaning of $\perp$ is derived, whereas the partial descriptions in (54b) and (54c) specify the constraints on the contributions of the quantifiers *every farmer* and *a donkey*, respectively.[56]

(54)

[56]The structural relations between elements of the derivation are of course very similar to those that would be assigned to the sentence in UDRT (Reyle, 1993), whose semantics is, however, rather different.


syn : [cat : s]
sem : [e|e:own(x1,x2), e < S]
us(=⊥)
syn : [cat : np]
sem : x1
a.
unp1
syn : [cat : vp]
sem : λy[e|e:own(y,x2), e < S]
uvp
phon : "owns"
syn : [cat : tv]
sem : λxλy[e|e:own(y,x), e < S]
uv
syn : [cat : np]
sem : x2
unp2
syn : [cat : s]
us1
b.
syn : [cat : np]
sem : λp[x1|farmer(x1)] ⇒ p
unp1
syn : [cat : s]
us1
phon : "every"
syn : [cat : det]
sem : λPλp[x1|]; P(x1) ⇒ p
udet1
phon : "farmer"
syn : [cat : n]
sem : λx[|farmer(x)]
un1
∧ ⊥ ◁* us1
∧ us1 ◁* ⊤

c.



$\wedge \perp \lhd^* u_{s_2}$
$\wedge\, u_{s^2} \lhd^* \top$

This partial interpretation is just what one gets from the lexicon under the assumptions about the lexical entries for noun phrases that we have discussed above and the assumptions about syntax and semantic interpretation discussed in previous sections.

6.2 Scope Disambiguation

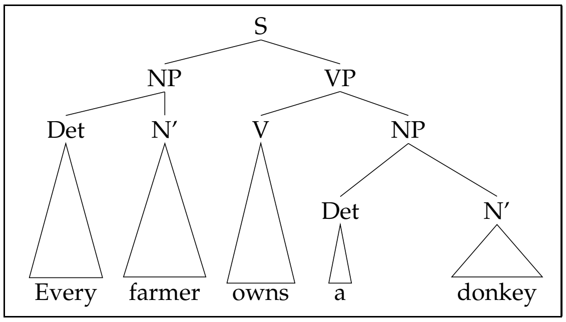
In this section we reformulate in terms of the theory of scope underspecification proposed here the theory of scope assignment presented in (Poesio, 1994)–which, in turn, was motivated by the results from the psycholinguistics literature discussed in Section §2 (especially (Kurtzman and MacDonald, 1993)) and by a corpus analysis of the scoping preferences of modals, definite descriptions, and indefinites in the TRAINS corpus of spoken dialogues collected at the University of Rochester (Gross *et al.*, 1993).

The theory in (Poesio, 1994) is based on the following additional hypotheses. First of all, scope assignment, much as the other types of semantic processing discussed in previous sections of this paper, can be formulated as a defeasible reasoning process involving defaults of different priorities. This hypothesis was made to account for the two fundamental results of Kurtzman and MacDonald: the existence of strong scopal preferences in (single clause) active sentences, and the disappearance of such scope preferences in passive sentences.

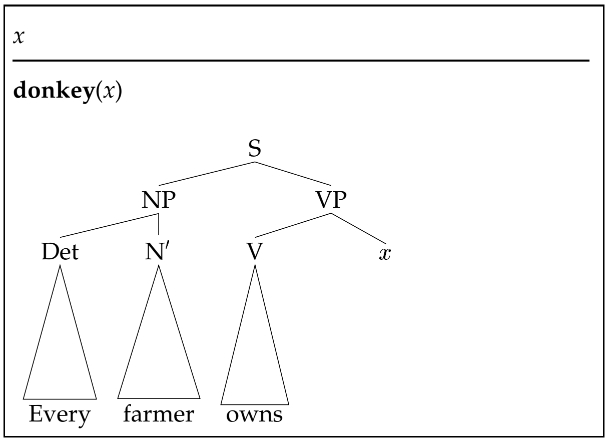
Secondly, it was assumed that scope disambiguation is based on rewriting operations similar to those proposed in DRT (Kamp and Reyle, 1993)–

i.e., operations that determine the scope of an operator relative to the other scope-taking elements in a sentence–rather than on operations that determine the relative scope of two operators (or their indices), as hypothesized in (Poesio, 1991; Alshawi and Crouch, 1992; Reyle, 1993). An example of the rewriting operations of DRT is the rewrite rule for indefinites, that determines the relative scope of the indefinite NP *a donkey* with respect to the rest of the interpretation of an utterance of *Every farmer owns a donkey*. This operation, according to Kamp and Reyle, takes as input a configuration like (55) and produces as output the configuration in (56).

(55)



(56)



Notice that these rewriting operations operate 'outside in' –which, as noted in (Lewin, 1990), eliminates the problems with 'inside out' operations of scope assignment which require a free variable constraints (Cooper, 1983; Hobbs and Shieber, 1987). However, rewriting operations such as these, if left unchecked, would simply start generating all possible readings of a quantified sentence, which is exactly what Kamp and Reyle want but does not give us a plausible theory of scope assignment. Poesio added therefore the additional stipulation that rewriting operations have to be *licensed*. The main licensing condition is that a rewriting rule can only apply to an operator after its context-dependent aspects have been determined. For example, the rewriting rule for quantifiers is only licensed after the domain of quantification of the quantifier has been determined, whereas the rewriting rule for modals only applies after the modal base (Kratzer, 1977) of the modal has been identified.

This licensing condition predicts that, in general, operators which have a contextual element that can be easily resolved in a given context will take scope over operators without a domain restriction (e.g., 'weak' indefinites) or operators whose domain restriction cannot be resolved on the basis of the existing context, and which therefore require some sort of presupposition accommodation; for example, modals will take scope over indefinites and negation. This predictions are borne out in the TRAINS corpus: e.g., the preferred reading of *We should not send an engine to Elmira* is 'all situations consistent with the plan are situations in which we do not send an engine to Elmira'.

According to this theory, then, the main task in developing a theory of scope disambiguation is to explain which operators depend on context for their interpretation, and to formulate theories about how these context-dependent aspects are resolved, to be formalized in terms of default inference rules. For reasons of space, we will not be able to discuss here these aspects of the theory, which take the largest part of Poesio (1994); instead, we will concentrate on the formalization of rewriting rules in terms of the theory of underspecification that we have introduced, which is different from the one assumed in (Poesio, 1994).

The hypothesis we pursue is that the 'rewriting operations' of Kamp and Reyle can be formulated in terms of ADJUNCTION in Tree Adjoining Grammar (Joshi, 1987), with the additional restriction that the adjunction always takes place in the same position, the one indicated by $\perp$ in Example (54). For example, the 'rewriting operation' that partially disambiguates (54) by assigning to *every farmer* wide scope consists of the adjoining in Figure 5.

Adjunction in Compositional DRT can be viewed as a combination of updates to the value of a discourse marker, here called A, together with a final operation of 'closure'. The current adjunction 'range' of an utterance is specified by the discourse markers A and $\perp$. Each time that a new operator gets adjoined, the utterance that immediately dominates it (u_{s^1}, in the example we have been discussing) gets identified with the previous value of A, whereas A gets updated to have as value the utterance that specifies the scope of this operator (u_{s_1}). For example, the steps involved in the disambiguation of (54) that assigns wider scope to *every farmer* are as follows:

1. $\perp \triangleleft^* \top; \mathrm{A} = \top$

2. $u_{s^1} \leftarrow \mathrm{A}; \mathrm{A} \leftarrow u_{s_1}$
   This adjunction (which must be independently licensed) assigns wide scope to *every farmer*.

3. $u_{s^2} \leftarrow \mathrm{A}; \mathrm{A} \leftarrow u_{s_2}$
   This adjunction assigns narrow scope to *a donkey*.

4. $\perp\leftarrow \mathrm{A}$
   This final update 'closes' the interpretation of the utterance, resulting in the interpretation in Figure 4.

7 Defeasible Reasoning and Ambiguity

In the previous sections we have seen how a number of disambiguation processes including lexical access, pronoun interpretation, and scope determination should be formalized in terms of defeasible reasoning. In this section we are going to generalize this and propose that all aspects of utterance processing, including parsing and semantic composition, should be formulated in terms of defeasible inference; and that this gives us not only a better account of these processes, but also a new perspective on what ambiguity is all about.

7.1 Parsing and Defaults

Just like lexical access can be viewed as a process of (prioritized) default inference, so can other aspects of interpretation–in particular, parsing and semantic composition. In Section §3.2 we did not specify how to interpret operators such as Left Attachment; now that we have introduced Prioritized Default Logic, we will propose that these operators are PDL default inference rules, as well.

Parsing Operations

That parsing is a defeasible reasoning process is by now a fairly established claim in the literature (MacDonald *et al.*, 1994; Stolcke, 1994; Jurafsky, 1996). This characterization is consistent with the fact that, whatever their interpretation as a characterization of grammatical competence, from the point of view of interpretation Left Attachment and other operators behave just like defaults: they suggest hypotheses–e.g., that the VP *raced past the barn* should be interpreted as constituting a sentence with the NP *the horse* in *the horse raced past the barn fell*–that can be defeated by further evidence with higher priority. And just as in the case of lexical entries, the defaults encoding grammatical performance have priorities which are used to resolve ambiguities.[57] A possible PDL formalization of left attachment is shown below.

$$\frac{\begin{array}{l}[|\ \mathbf{syn}(u_1)\ \mathbf{is}\ \begin{bmatrix}\mathsf{cat : A}\end{bmatrix}, \mathbf{syn}(u_2)\ \mathbf{is}\ \begin{bmatrix}\mathsf{cat : A}\end{bmatrix}, \\ \mathbf{syn}(u_3)\ \mathbf{is}\ \begin{bmatrix}\mathsf{cat : D}\end{bmatrix}, u_2 \lhd u_3, u_1 \unlhd u_2]\end{array} \quad : \ [|u_1\ \mathbf{is}\ u_2]}{[u_1\ \mathbf{is}\ u_2]}\ \mathbf{LA}$$

Semantic Composition

Once all word utterances have been given a lexical interpretation and the constituent structure in which they participate has been identified, the two semantic composition rules discussed in Section §3.2 apply to compositionally derive the meaning of phrasal utterances. In this section, we would like to examine this process more closely, and ask ourselves the question: what is the nature of these rules? Are they defeasible, like lexical rules and syntactic composition operators, or not?

Phenomena such as metonymy or 'transfers of meaning' Nunberg (1978, 1995) suggest that semantic composition rules may also behave like defaults. Nunberg (1995) identifies two types of metonymy: DEFERRED INDEXICAL REFERENCE and PREDICATE TRANSFER. These two types of metonymy are illustrated by the two utterances in (57), to be imagined uttered by a customer handing his key to an attendant at a parking lot:

(57) a. This is parked out back.
b. I am parked out back.

Deferred indexical reference is the case of metonymy in which an NP refers indirectly; Nunberg argues that in (57a), *this* refers to the car rather than to the key, as shown by the fact that in languages in which objects may have different genders, the gender of the car, not of the key, would used in this case. Predicate transfer, illustrated by (57b), is the analogous case for predicates: Nunberg argues that in this example, *am parked out back* is not

[57] Recent work on grammar inspired by Optimality Theory (Prince and Smolensky, 1993) can be interpreted as suggesting that interpreting grammatical rules as defaults with priorities is also important for a theory of competence. See e.g., (Bard *et al.*, 1996; Keller, 1996).

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:

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

Nunberg argues that predicate transfer is '. . . a phrasal phenomenon that works in concert with the process of semantic composition' and is subject to the same constraints; e.g., composition has to apply in a certain order. This is shown by the fact that although predicate transfer can apply in (59a) to get a 'disease' reading out of 'virus', and although predicate transfer can apply after 'virus' has been modified so that (59a) can be interpreted as referring, for example, to 'a widely studied disease caused by a virus native to Peru', that NP does not have the reading in (59b), obtained by applying 'Peruvian' to the result of applying 'widely studied' to 'virus':

(59) a. *a widely studied Peruvian virus*
b. 'a disease endemic to Peru that is caused by a widely studied virus'

Nunberg's description of semantic composition makes it sound very much like 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 . . .

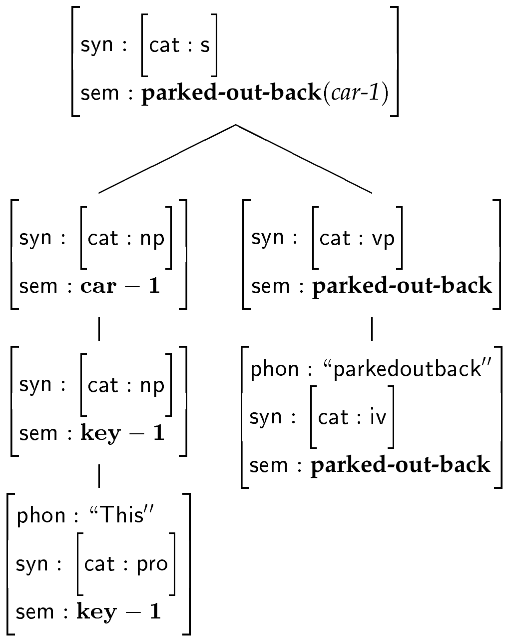
There are two possible strategies to account for Nunberg's observations, both of which involve making semantic composition into a defeasible process. Perhaps the simplest strategy is to hypothesize that there is more than one way of obtaining the meaning of an utterance of type s from the meanings of its constituents of type np and vp; rules such as those discussed in Section §3.2–e.g., applying the direct meaning of the np to the direct meaning of the vp–are only one way of doing this, that can be overridden by more specific strategies. We could hypothesize that at least two defaults can apply to derive the meaning of a node from the meanings of its constituents: a low-priority one, **BINARY-SEMANTIC-COMPOSITION**, doing what the semantic composition rule **Binary Semantic Composition** described in Section §3.2 does; and a higher-priority one, **PT-BINARY-SEMANTIC-COMPOSITION**, 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 (58)):

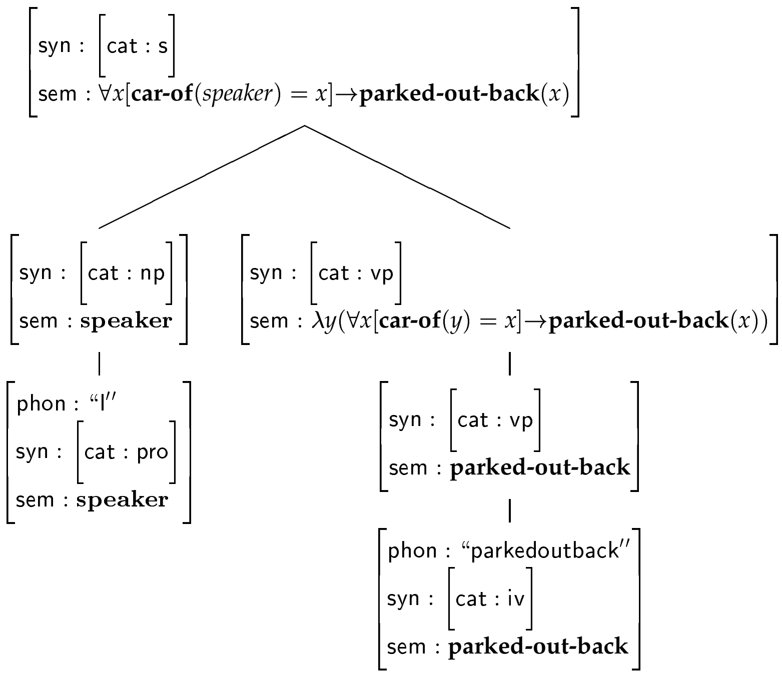
$$\frac{\begin{array}{ll} u_1 \lhd u_3 \wedge u_2 \lhd u_3 \wedge & \\ u_1 \rightsquigarrow \varphi_{\langle\alpha,\beta\rangle} \wedge u_2 \rightsquigarrow \psi\alpha \wedge & : u_3 \rightsquigarrow \Upsilon(\psi) \\ \textbf{COMPLETE}_3(u_3,u_1,u_2) \wedge & \\ \exists \text{ g g}(\varphi) = \Upsilon & \end{array}}{u_3 \rightsquigarrow \Upsilon(\psi)} \textbf{PT-BINARY-SEMANTIC-COMPOSITION}$$

An alternative formalization, perhaps closest to what Nunberg has in mind, involves modifying our assumptions about the type of constituent structure that gets constructed during interpretation. Until now, we have been implicitly assuming that deriving the conventional meaning involves constructing a hierarchical structure of utterances analogous to what used to be called s-structure. An alternative way of interpreting Nunberg's suggestions is to look at them as a proposal that what gets constructed during interpretation is a dominance hierarchy that is more like what is called a 'derivation tree' in Montague grammar, or LF in a generative framework. If we adopt this perspective, the rules allowing for the derivation of 'transferred' meanings can be seen as rules allowing to make an additional and optional derivation step, provided that it is consistent to do so. The (much simplified) derivation of the conventional meaning of *This is parked out back* would then be as in (60a), whereas the derivation of the conventional meaning of *I am parked out back* would be as in (60b).

(60) a.



b.



We have no way of resolving the question of which of these two hypothesis is correct for the moment; however, as we seen in Section §6, a number of recent approaches to scopal underspecification assume that deriving the conventional meaning of an utterance does involve constructing something like the LF of that utterance.

7.2 Extensions and Precisifications

We can now return to the question raised at the beginning of this section: how is the intuition that ambiguity has to do with 'multiplicity of senses' captured by the theory presented here? The answer suggested in (Poesio, 1996b) is that a theory of defeasible reasoning like Prioritized Default Logic provides us with a notion of precisification[58] to which is perfectly adequate for our purposes: namely, the notion of extension. More precisely, we can define precisifications as follows. Let α be an expression of the language, and let α' be its h-underspecified interpretation. (E.g., α could be *Kermit croaked*, whereas α' would be

$$[u|u{:}\mathbf{utter}(a,\text{“}\textit{Kermit croaked}\text{”})].)$$

Let context C be the PDL theory $\langle$D,W$\rangle$. Then,

Definition 5 *The* PRECISIFICATIONS *of α in C are the extensions of the* PDL *theory $\langle D,W \cup \alpha' \rangle$.*

We can then define ambiguity as follows:

Definition 6 *Expression α is* AMBIGUOUS WITH RESPECT TO CONTEXT C, *where C = $\langle D,W \rangle$, if $\langle D,W;\ [u|u{:}\mathbf{utter}(A,\alpha)] \rangle$ has more than one extension.*

More specifically, *semantic* ambiguity has to do with differences in the value of the function $\leadsto$:

Definition 7 *Expression α is* SEMANTICALLY AMBIGUOUS WITH RESPECT TO CONTEXT C *if it is ambiguous with respect to C, and if the extensions of $\langle D,W;\ [u|u{:}\mathbf{utter}(A,\alpha)] \rangle$ include extensions E and E$'$ such that $E \models u \leadsto m$, $E' \models u \leadsto m'$, with $m \neq m'$.*

In other words, by combining a theory of defeasible reasoning with the properties discussed in this section, with the theory of underspecification introduced in the previous sections, we get a theory of ambiguity in which the notion of precisification is not used to define the semantics of the logical

[58] We refer to the disambiguations of an expression by using the term 'precisification' proposed by Pinkal (1995).

language, unlike in most early theories of underspecification (Alshawi and Crouch, 1992; Poesio, 1991, 1996a; Reyle, 1993). In the theory presented here, the notion of underspecification is formalized using Compositional DRT –that is, a fairly traditional logical theory. This account of underspecification and ambiguity has the advantage that it ties in easily with current work in semantics: we do not need a new logic for characterizing ambiguous expressions.

On the other hand, some may think that the approach has the drawback that we get a context-dependent notion of ambiguity: in our theory, expressions are always ambiguous in a relative sense, but with respect to a context C (i.e., a specific default theory). By contrast, (semantic) ambiguity is traditionally defined in absolute terms. We believe however that the precisifications-as-extensions theory has more intuitive appeal than the precisifications-as-semantic-competence theory. For one thing, it fits better with the observation that most people cannot explain what a word means unless they are given some 'context'—the rest of the sentence in which the word occurs. In addition, we would argue that the precisifications-as-extensions theory also gives us a simple explanation of why linguists tend to find sentences more ambiguous than non-linguists—their 'context' would always include more linguistic defaults.[59]

8 Do Interpretations Remain Underspecified? Evidence from Corpora

Perhaps the most intriguing issue raised by some of the original work on scopal underspecification was the possibility that the conventional meaning of utterances may not always be completely identified at the end. Given that in most recent theories of underspecification the underspecified interpretation of an utterance is a first-class logical object with a proper semantics, it is possible to use this interpretation as the basis for some types of inference even without achieving a complete disambiguation; this makes complete disambiguation optional (and pragmatically controlled), rather than mandatory. This idea was a radical departure from the traditional assumptions of mainstream semantic theory.

One could also hypothesize 'mixed' degrees of disambiguation for different phenomena. For example, one of the central claims of Construal The-

[59]The idea that ambiguity may be context-dependent was also suggested by Pulman (1994).

ory (Frazier and Clifton, 1997) is that certain aspects of syntactic interpretation are completely determined, whereas others (such as adjunct attachment) are left underspecified. In a similar fashion, one may ask whether the interpretation of an utterance containing a pronoun such as *he left* in a context in which no suitable antecedent is available, or in cases like those discussed by Garrod and Sanford, remains h-underspecified, or if instead in that context humans produce a weak interpretation such as 'there is a human individual that left'.

Ultimately, this question can only be resolved by psycholinguistic research; but at the light of the work on underspecification of the past few years it is already possible to attempt some preliminary conclusions. Throughout this paper we have discussed, for the most part, evidence that an utterance's conventional meaning is either immediately identified, or identified by the end of the sentence; and this conventional meaning is generally of the type assumed in conventional semantic theory. Indeed, these are the only possibilities according to Frazier and Rayner's IPIH. The exception to this are the results reported by Garrod and Sanford; in those cases, it does seem like certain expressions are left uninterpreted. In the case of scope, the evidence is less clear: we know from examples like (9b) that we can have scopal garden paths, and the experiments of Kurtzman and MacDonald indicate that subjects do have preferences (Kurtzman and MacDonald, 1993; Poesio, 1994), but it's not clear how more complex sentences with more than two quantifiers, like (7), are interpreted.

In this section we want to look at this question by using a different type of evidence. Over the past few years we have been uncovering in the corpora we have been studying evidence that certain expressions may not have a 'unique' disambiguation in the given text, and that subjects may be satisfied with a partial resolution. One example are dialogues like the following (dialogue d91-3.1 from the TRAINS corpus at the University of Rochester, `http://www.cs.rochester.edu/research/speech/dialogues.html`):

(61)
```
3.1     M: can we .. kindly hook up
3.2      : uh
3.3      : engine E2 to the boxcar at .. Elmira
4.1     S: ok
5.1     M: +and+ send it to Corning
5.2      : as soon as possible please
6.1     S: okay
         [2sec]
7.1     M: do let me know when it gets there
8.1     S: okay it'll /
```

```
8.2        : it should get there at 2 AM
9.1      M: great
9.2        : uh can you give the
9.3        : manager at Corning instructions that
9.4        : as soon as it arrives
9.5        : it should be filled with oranges
10.1     S: okay
10.2       : then we can get that filled
```

In this example, it's not clear whether the pronoun *it* in 5.1 refers to *engine E2* or to *the boxcar at Elmira*, or if indeed that matters; it's only at utterance 9.5 that we get evidence that *it* probably referred to *the boxcar at Elmira*, since it is only boxcars that can be filled with oranges.

Other cases of apparently ambiguous references include certain references to surface areas discussed in (Poesio and Vieira, 1998) and references to abstract objects such as plans discussed in (Poesio, 1994). In trying to understand what happens in these cases, we have to address two questions: why the writer / speaker can afford leaving the interpretation partially open (is he/she simply being sloppy?), and what does the reader/ listener do. We briefly discuss a few other cases; more evidence, and our preliminary hypotheses, are discussed in (Poesio and Reyle, 1999).

Mereology Cases

'Sloppy' references such as the pronoun in (61) are fairly common in the TRAINS dialogues: here are two additional examples (from d91-2.2 and d91-2.1, respectively).[60] In (62), the question is whether the pronoun *it* in 44.5 refers to *the engine from Avon* or *the boxcar*.

(62)
```
42.1 S: if we took the engine from Avon
43.1 M:                         +to Bath+ apparently
44.1 S: if we took it to Bath
44.2  : that would take 8 hours
44.3  : the same amount
44.4  : to get the boxcar and come back
44.5  : if we took it to Dansville
44.6  : take 6 hours to
44.7  : get the boxcar and come back
```

(63)
```
18.1 S: um
```

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

```
18.2  : y / you could take engine E1
18.3  : and get
18.4  : a boxcar from Bath to Corning in 6 hours
18.5  : or you could take engine E2 or .. E3
18.6  : and get a boxcar to Corning
18.7  : in 2 hours
 [4sec]
19.1 M: okay
19.2  : so why don't we do that
19.3  : take one of the engines from Elmira
19.4  : and the boxcar
19.5  : and take it to Corning
20.1 S: +okay+
20.2  : so we'll say E2
20.3  : and the boxcar
20.4  : take it to Corning
```

In the TRAINS corpus we also find a second pattern, illustrated by (64), in which the pronoun *that* in 27.4 may refer either to the orange juice or to the tanker car in which the orange juice has been loaded:

(64)
```
26.1 S: okay
27.1 M: so then we'll
27.2  : ... we'll be in a position to
27.3  : load the orange juice into the tanker car
27.4  : ... and send that off
```

Intuitively, both in situations involving attaching two objects together and in situations involving loading objects into other objects, the two available interpretations of the pronoun are equivalent as far as the plan is concerned: if one object gets moved, the other one must be moved as well, because they are part of the same mereological structure. We will write X $\sim$ Y to indicate that X is equivalent to Y for the purpose of the plan: e.g., in the case of (61), we will write $e \sim b$ to say that from the point of view of the plan, the interpretation of *it* in which it refers to engine E2 and that in which it refers to the boxcar are equivalent. We will use Link's notation $e \oplus b$ to indicate the object that has e and b as subparts, and $e \lhd t$ to say that e is a mereological part of t.

What's interesting about these examples is that in all of them the speaker is taking a risk, yet the listener does not signal a problem in understanding. One question to be asked, then, is whether the speaker is simply being sloppy, or he/she has done what Hobbs (1985) would call 'collapsing a complex theory in one of coarser granularity'–i.e., he/she is aware that there are two possible interpretations, but is also aware that the two interpretations are equivalent. This question is unfortunately difficult to answer.

A second question, and one perhaps more amenable to empirical testing, is why isn't the listener complaining. The space of possible answers is as follows:

1. The listener doesn't even attempt to interpret the pronoun, and keeps for it an h-underspecified interpretation. This hypothesis would be perhaps more plausible in the case of non-task-oriented dialogues such as those in the Switchboard corpus; less so in the case of task-oriented dialogues. Furthermore, the psychological evidence about anaphora discussed above suggests that listeners do interpret pronouns when the entity they refer to is in focus (as is the case in all examples here).[61]

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

   (a) The listener realizes that there are two possible interpretations. In this case, there are three more possibilities:

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

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

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

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

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

It's clear that the corpus does not give us enough information to choose among the hypotheses; it's also possible that different listeners adopt dif-

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

ferent strategies. A few controlled psychological experiments are needed to resolve this question.

References to spatial areas

In a recent empirical study of the interpretation of definite descriptions (Poesio and Vieira, 1998) it was reported that human subjects do not agree on the interpretation of definite descriptions such as *the area* in the following example:

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

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

In this example, again, we have an underlying mereological structure: both *parts of the factory* and *areas of the factory* are obviously included in the total area of the factory. However, unlike the cases discussed above, there is no obvious equivalence between the three different interpretations. Also, this is the case in which one can make the clearest argument that the object referred to - the area - is not the 'topic' of the text; to the extent that one can say that there is a topic in this text, it is most likely the factory. So, this is perhaps the example in which it is most likely that Garrod and Sanford's finding applies –i.e., the reader did not even attempt to construct an interpretation for the anaphoric expression.

9 Related Research

'Underspecified logical forms' were proposed early on in computational linguistics as a way of separating the task of grammar from that of the pragmatic component (Woods *et al.*, 1972; Schubert and Pelletier, 1982; Moore, 1981). However, these early proposals were not concerned with providing a real semantics for these interpretations, i.e., to specify which information the language processor had after producing these interpretations; nor with the mechanics of the disambiguation process, typically formulated in terms of heuristics (e.g., for scope, (Hurum, 1988; Moran, 1988)) and with little input from psychological research (admittedly, there wasn't much available at the time).

The first to formulate a theory of disambiguation in terms of inferences over underspecified interpretations with a proper semantics was Hobbs (Hobbs, 1982, 1983; Hobbs *et al.*, 1993). Hobbs characterized initial interpretations in terms of a very flat first-order logical representation (e.g., *Kermit croaked* would be represented as in (66)) and formalized disambiguation as a process of abductive inference.[62]

(66) $\exists w1, w2$ **Kermit**$(w1) \wedge$ **croaked**$(w2)$

Both the characterization of interpretations and the theory of disambiguation presented here have been influenced by Hobbs' proposals, which, however, has many problematic features. The main problem perhaps is that Hobbs proposes for most phenomena analyses which are radically different from those adopted in traditional semantic theory. Part of the reason for this is to avoid committing early on to a given interpretation; we hope to have showed that this goal can be achieved while maintaining more traditional analyses.[63]

Incremental theories of semantic processing have been proposed, among others, in (Haddock, 1988; Pereira and Pollack, 1991; Shieber and Johnson, 1993; Milward and Sturt, 1995). Shieber and Johnson (1993) propose a model of interpretation which is mainly concerned with accounting for incrementality effects on parsing, rather than for evidence about underspecification and about processes such as lexical disambiguation or pronoun resolution. Their model of parsing is also based on TAGs, like Sturt and Crocker's, but they tie syntactic information with semantic information in so-called ASYNCHRONOUS TAGS, except that the semantic components of asynchronous tags are much more 'representational' in nature than the semantic representation proposed here–i.e., these semantic interpretations are objects that can be manipulated by means of operations very similar to those that operate on syntactic representations.[64]

[62] Recently, 'flat' logical representations have also been adopted in VERBMOBIL (Bos *et al.*, 1994) and other projects involved in robust interpretation of dialogues.

[63] A second reason is Hobbs' desire to preserve a first-order representation; all we can say is that at the moment it is not at all clear that a first-order representation is enough to capture all the complexities of natural language semantics. See, however, (Hwang and Schubert, 1993).

[64] See also the discussion in Section §5 about Kamp and Reyle's proposals about under-

Jurafsky (1996) proposes a model of disambiguation motivated by much the same considerations as the model presented here, except for the data about underspecification, but that again concentrates on syntactic rather than semantic disambiguation. One of the main points of Jurafsky's paper is to argue for a probabilistic model of disambiguation. Although our model is not cast in probabilistic terms, we believe that our account of the evidence which is relevant to both papers is largely compatible with his.

10 Summary And Open Problems

10.1 A Summary

One contribution of this paper is our generalization of ideas about interpretation developed in work on syntactic underspecification to a model in which the language processor may also entertain semantically underspecified interpretations, while processing the input incrementally and generating hypotheses in parallel. The notion of interpretation we developed is centered around the idea, shared with sign-based theories, that the interpretations constructed by the language processor include the information that there are one or more 'language objects' that have to be processed – which, in our case, are utterances, a generalization of the notion of sign. Adopting this view of interpretation allowed us to explain how semantic interpretation can proceed incrementally while preserving compositional meaning construction, and to develop a theory of semantic underspecification that does not require a specialized logical formalism. We also attempted to clarify previous discussions among computational semanticists by drawing a distinction between two senses in which an interpretation can be underspecified: h-underspecification (the case in which the value of one of the attributes of an utterance is not specified by the interpretation) and p-underspecification (the case in which the interpretation specifies a provisional value, to be further refined by pragmatic reasoning after the interpretation of the whole utterance is completed).

The second ingredient of our theory of utterance processing is a very general characterization of disambiguation as a process of inference in Prioritized Default Logic. This part of the proposal is needed to provide a formal characterization of the notions of parallel hypothesis generation, of pruning, and of the situations in which the language processor still maintains multiple hypotheses even after preliminary pruning. We also claimed

specified anaphoric expressions.

that a theory of defeasible reasoning with priorities (whether statistical or nonmonotonic) is not just necessary to provide an account of disambiguation; it's a crucial ingredient of a theory of ambiguity.

We used our theory to give a formal account of the currently available evidence about semantic underspecification: this includes the results about lexical access by Frazier and Rayner (1990), and the results on pronoun interpretation by Corbett and Chang (1983); Gernsbacher (1989); Garrod and Sanford (1994). Our account of pronoun resolution involved a formalization of centering theory which is also novel and is in accord with the formal intuitions that stood behind the original formulation of the theory (Grosz *et al.*, 1995). We also proposed a preliminary account of some evidence about scoping, and studied corpus-based evidence suggesting that in some cases, the semantic interpretation of an utterance may not be completely determined.

One of the conclusions that can be drawn from this work is that even Frazier and Rayner's IPIH is still too strong. The language processor can maintain multiple interpretations in parallel at least until the end of the sentence; and if our account of Garrod and Sanford's results is correct, there may be utterances whose conventional meaning remains h-underspecified.

10.2 Open Problems

We think that an equally important contribution of this paper is our attempt to provide a systematic survey of currently available evidence on semantic underspecification. We mentioned that psychological research is needed both to get more information on scope assignment and to determine the form of the final interpretations in the cases in which there doesn't seem to be a preferred disambiguation of an utterance, discussed in Section §8.

Concerning the technical aspects of the proposal, perhaps the most urgent need is to find a way to connect theories of semantic disambiguation based on semantic theories motivated by linguistic evidence, such as the one presented here, with theories of semantic disambiguation based on considerations about learning, such as, for example, (Schütze, 1997). This may be less difficult than some people believe: contrary to common belief, logic formalisms are not inherently tied to symbolic assumptions (see, e.g., (Pinkal, 1995) for how to build a logic formalism without assuming a fixed denotation for the logical constants); only 'representational' formalisms are. Nevertheless, doing this will be a major undertaking. We also need better theories of the dynamic re-ranking of priorities during interpretation, to account for the results about lexical disambiguation and pronoun resolution reported in (Simpson, 1994; Garrod and Sanford, 1994).

Acknowledgments

Special thanks to Patrick Sturt for detailed comments on an earlier version of this paper and to Robin Cooper, Richard Crouch, Kees van Deemter, Jonathan Ginzburg, Janet Hitzeman, Reinhard Muskens, Manfred Pinkal, Steve Pulman, Uwe Reyle, Rosemary Stevenson, David Traum, and Roberto Zamparelli for numerous comments and suggestions about these issues over the last several years. We also wish to thank James Allen, Nick Asher, Ellen Bard, Matt Crocker, Tim Fernando, Lyn Frazier, Graham Katz, Alex Lascarides, David Milward, Hannes Rieser, Len Schubert, and Alessandro Zucchi. This work was supported by an Advanced Research Fellowship from the EPSRC.

A Compositional DRT

A.1 Basic Ideas

Compositional DRT (Muskens, 1995b, 1994) merges the familiar technical tools from Intensional Logic with technical ideas from Discourse Representation Theory (DRT) (Kamp and Reyle, 1993), including DRT's treatment of events (that we will use to assign an interpretation to the expressions describing utterances seen in the previous sections) and DRTs' treatment of context change and anaphora.[65]

We recall that the fundamental idea of dynamic theories is that sentences are seen as transitions among STATES: that is, a sentence like (67)

(67) *A man left*

is thought of as specifying a transition from initial states in which no DISCOURSE ENTITY for 'a man' is available to ones in which it is and in which, furthermore, this discourse entity (let's call it x) satisfies the conditions imposed by the sentence (i.e., that x is a man and that it left). The sentence that follows (67) in a discourse will, in turn, specify a transition from any state which results from the transition specified by (67) to new states in which

[65] The theory of grammar that we propose can be seen as a generalization of dynamic theories such as DRT, in that natural language expressions specify context updates; our characterization of the results of interpreting anaphoric expressions is also based on DRT's treatment.

additional conditions are imposed on x and new discourse entities may be introduced.

In DRT, states are represented by Discourse Representation Structures (DRS), and the state change resulting from a sentence is specified by means of algorithms that modify an existing DRS by adding new discourse entities and / or conditions. Compositional DRT is based on the rather different view of dynamics adopted in (Heim, 1982; Barwise, 1987; Rooth, 1987; Groenendijk and Stokhof, 1991), according to which states are similar to variable assignments, and sentences are viewed as relations among states. In Compositional DRT, DRSs are expressions that encode state changes; sentences are translated into DRSs, and the meaning of a text is obtained by 'concatenating' the DRSs associated with its constituent sentences, which allows for a compositional procedure of meaning construction. In symbols, and using Muskens' linear notation for DRSs according to which DRSs are represented by expressions of the form

$$[\mathrm{u}_1, \ldots, \mathrm{u}_n \mid \varphi_1, \ldots, \varphi_m]$$

the interpretation of (67) as a relation between assignments can be specified in semi-formal terms as in (68).

(68) $[\![[x|\mathbf{man}(x),\mathbf{left}(x)]]\!] = \{\langle \mathrm{i,j} \rangle \mid$ i and j are states, j differs from i at most over x, and the values assigned by j to x satisfy $[\![\mathbf{man}(x)]\!]$, $[\![\mathbf{left}(x)]\!]\}$

In (68), the denotation of a DRS is specified as a relation between states, i.e., a set of pairs $\langle \mathrm{i,j} \rangle$ of states. The interpretation of a discourse is constructed by concatenating the interpretations of each sentence by means of the concatenation relation among DRSs ';'. If K and K′ are DRSs, K;K′ specifies a transition from a set of initial states to a new set of states that satisfy both the constraints imposed by K and those imposed by K′.

The central idea of Compositional DRT is to define the main expressions of DRT, such as DRSs, as expressions in type theory. This is done by adding to the primitive types e ('entities') and t ('truth values') of Intensional Logic, the two primitive types s–the type of STATES–and π_e– the type of DISCOURSE ENTITIES (of type e).[66]

Constants of type π_e are used to represents discourse entities, as follows. In order to allow these constants to occur as arguments of predicates, a constant $\mathbf{V}_e$ of type $\langle \pi_e, \langle s,e \rangle \rangle$ is introduced, denoting a function from discourse entities and states to entities; this function specifies the object associated with discourse entity d at state i. In this way, we can write expressions

[66] As we will see shortly, in Compositional DRT there are other types of discourse entities besides π_e.

such as **man**($\mathbf{V}(x,j)$), meaning that the value of discourse entity x at state j is a man. (In fact, in what follows instead of writing $\mathbf{V}(x,j)$ every time, we will simply write the discourse entity as an argument of the predicate, as done by Muskens: e.g., we will write **man**(x) instead of **man**($\mathbf{V}(x,j)$).) A second constant $\mathbf{ST}_e$ of type $\langle \pi_e, t\rangle$ is also defined; $\mathbf{ST}_e(x)$ is true iff x is a constant of type π_e.

State changes are specified by means of an 'update relation' $i[u]j$, which holds between states i and j if j differs from i at most over the value assigned to discourse entity u of type π_α:

- $i[u]j$ is short for (i) $\forall v\ (\mathbf{ST}_\alpha(v) \wedge\ u \neq v) \rightarrow (\mathbf{V}_\alpha(v,i) = \mathbf{V}_\alpha(v,j))$ and (ii) the conjunction of all $\forall v_{\pi_\beta} \mathbf{ST}_\beta(v) \rightarrow (\mathbf{V}_\beta(v,i) = \mathbf{V}_\beta(v,j))$ for all $\beta \in \Theta - \{\alpha\}$, where Θ is the set of types of objects which can be values for discourse entities (to be specified below).
- $i[]j$ is short for $\forall v\ \mathbf{V}_\alpha(v,i) = \mathbf{V}_\alpha(v,j)$ for $\alpha \in \Theta$.

The definition of $i[u]j$ can be generalized to an arbitrary number of discourse entities by defining $i[u_1, \ldots, u_n]$ by induction as an abbreviation of the formula $\exists k(i[u_1]k \wedge\ k[u_2, \ldots, u_n]j)$.

The type-theoretic interpretation of the constructs of DRT is specified as follows: let K and K′ be DRSs, i and j variables ranging over states, and $\varphi_1, \ldots, \varphi_m$ expressions of type $\langle s,t\rangle$. Then DRT's simple and complex conditions (Kamp and Reyle, 1993), as well as the concatenation of DRSs K;K′, are defined as follows:

	$\mathbf{R}\{\tau_1, \ldots, \tau_n\}$	is short for	$\lambda i \mathbf{R}(\tau_1) \ldots (\tau_n)$
	τ_1 **is** τ_2		$\lambda i (\tau_1) = (\tau_2)$
	not(K)		$\lambda i.\ \neg \exists j K(i)(j)$
(69)	K **or** K'		$\lambda i \exists j K(i)(j) \vee K'(i)(j)$
	$K \Rightarrow K'$		$\lambda i \forall j K(i)(j) \rightarrow \exists k K'(j)(k)$
	$[\mathrm{u}_1, \ldots, \mathrm{u}_n \mid \varphi_1, \ldots, \varphi_m]$		$\lambda i.\ \lambda j.\ i[\mathrm{u}_1, \ldots, \mathrm{u}_n]j \wedge\ \varphi_1(j), \ldots, \varphi_m(j)$
	$K ; K'$		$\lambda i.\ \lambda j.\ \exists k K(i)(k) \wedge\ K'(k)(j)$

where the expressions on the left-hand side are expressions of the DRT language together with concatenation K;K′, and those on the right-hand side are expressions of the standard type-theoretical language defined below. Conditions are defined as functions from states to truth values; DRSs as functions from pairs of states onto truth values. A DRS K with discourse entities $u_1 \ldots u_n$ and conditions $\varphi_1 \ldots \varphi_m$ is true at state i iff there is a state j such that $\langle i,j\rangle$ is in the denotation of K, i.e., such that j agrees with i over all discourse entities other than $\mathrm{u}_1 \ldots \mathrm{u}_n$, and all of φ_l hold at j. For example, if we ignore tense, the DRS representation of *A man left* in (70b) is interpreted as in (70c).

(70) a. *A man left*

b. $[x|\ \mathbf{man}(x), \mathbf{left}(x)]$

c. $\lambda i \lambda j [i[x]j \wedge \mathbf{man}(\mathbf{V}(x,j)) \wedge \mathbf{left}(\mathbf{V}(x,j))]$

The Unselective Binding Lemma asserts that these definitions yield the right semantics for DRSs (e.g., they assign existential force to a DRS like (70b)); his Merging Lemma ensures that the DRS for a discourse can be composed piecemeal from the DRSs of single sentences and ';'.

The concepts introduced so far are sufficient to reformulate the treatment of (singular) nominal anaphora of Kamp and Reyle (1993); additional machinery is introduced in Compositional DRT to reproduce the treatment of tense and event anaphora. The treatment of tense adopted in Muskens (1995b) is Davidsonian (Davidson, 1967), in the sense that each predicate (both nominal and verbal) is assumed to have an additional argument for the EVENTUALITY of that state or event. Eventualities, as well as temporal intervals, are treated as first-class objects: in addition to the types e, t, s and π discussed so far, there are in Compositional DRT a type ϵ for events and a type τ for temporal intervals. Periods of time are viewed as segments of a Euclidean straight line; eventualities are allowed to branch in the future. Each eventuality is associated with a period at which it takes place. (The axioms for eventualities and time intervals are given below.)

As in (Kamp and Reyle, 1993), it is assumed that temporal expressions introduce new discourse entities. For example, *A man left* introduces into discourse an event of a man leaving, which can subsequently be referred to (as in, e.g., *that happened at 5pm*). In DRT notation, this sentence is translated as (71b), where S is a discourse entity denoting the speech time; this expression is equivalent to (71c) in type-theoretical terms.[67]

(71) a. *A man left*

b. $[x,e_1|\ \mathbf{man}(x), \mathbf{left}(x,e_1), e_1 < S]$

c. $\lambda i \lambda j i[x, e_1]j \wedge \mathbf{man}(x) \wedge \mathbf{left}(x, e_1) \wedge e_1 < S$

In what follows, we will use the expression e:$\mathbf{p}(\overline{x})$ as a shorthand for what Muskens would write as $\mathbf{p}(\overline{x}, e)$, in accord with the DRT notation used in (Kamp and Reyle, 1993): e.g., we will write e_1:$\mathbf{left}(x)$ instead of $\mathbf{left}(x,e_1)$.

The discourse entities holding values of type eventuality and time interval are handled in the same way as discourse entities holding values of type e: they are represented as constants of types π_τ and π_ϵ, whose values

[67]We have simplified matters considerably with respect to Muskens' treatment in Muskens (1995b); in particular, we have ignored the issue of the referentiality of times and reference time shifting.

at a given state are specified by constants $\mathbf{V}_\tau$ and $\mathbf{V}_\epsilon$, of type $\langle\pi_\epsilon,\langle s,\epsilon\rangle\rangle$ and $\langle\pi_\tau,\langle s,\tau\rangle\rangle$, respectively. Overall, the set Θ of types that can be assigned as values of discourse entities in (Muskens, 1995b) is $\Theta = \{e,\epsilon,\tau,w\}$, where w is the type of worlds (i.e., Intensional Logic's type s).

A.2 More Details

In this section we give more details about Compositional DRT; in particular, about its axiomatization. This latter is important for the formalization of the disambiguation process in Section §4, but can be skipped by those readers not interested in the details of the formalization of reasoning we are using.

The set of types of Compositional DRT is the minimal set such that

1. e, t, s, ϵ, τ, w (for worlds), π_e, π_ϵ, π_τ and π_w are types;
2. if α and β are types, then so is $\langle\alpha,\beta\rangle$.

For each type an infinite set of variables and constants of that type is assumed. The 'basic terms' of the language are defined in the standard way, as follows:

1. Every constant or variable of any type is a term of that type;
2. If φ and ψ are terms of type t (FORMULAE), then $(\neg\varphi)$ and $(\varphi\wedge\psi)$ are formulae;
3. If φ is a formula and x is a variable of any type, then $(\forall\ x\ \varphi)$ is a formula;
4. If A is a term of type $\langle\alpha,\beta\rangle$ and B is a term of type α, then (A B) is a term of type β;
5. If A is a term of type β and x is a variable of type α, then $\lambda\, x.$ A is a term of type $\langle\alpha,\beta\rangle$;
6. If A and B are terms of type α, then (A = B) is a formula.

The set of non-logical constants includes:

1. the constants $\mathbf{V}_e$ of type $\langle\pi_e,\langle s,e\rangle\rangle$, $\mathbf{V}_\epsilon$ of type $\langle\pi_\epsilon,\langle s,\epsilon\rangle\rangle$, $\mathbf{V}_\tau$ of type $\langle\pi_\tau,\langle s,\tau\rangle\rangle$, and $\mathbf{V}_w$ of type $\langle\pi_w,\langle s,w\rangle\rangle$;
2. the constants $\mathbf{ST}_e$, $\mathbf{ST}_\epsilon$, $\mathbf{ST}_\tau$, and $\mathbf{ST}_w$, defined as above;

3. the constant $<$ of type $\langle\tau,\langle\tau,t\rangle\rangle$ (see below);

4. the constant ϑ of type $\langle\epsilon,\tau\rangle$ (see below).

To these expressions we add the expressions of the DRT language, defined as in (69). The semantics of these expressions is defined in the usual fashion; most interesting constraints on the model are imposed by means of the axioms below.

One of the essential properties of Compositional DRT for our purposes is that it has a sound and complete proof theory; this is a necessary prerequisite for the formulation of reasoning with underspecified interpretations that we will present in the following sections. Sound and complete axiomatizations for type theory (with respect to generalized models) are given by Henkin (1950) and Andrews (1986); the formulation we give below is Carpenter's formulation of Andrews' axioms (Carpenter, 1998). These basic axioms consist of all the instances of the following schemas:

BA1 $P(true) \wedge P(false) = \forall x P(x)$

BA2 $(x = y) \rightarrow (P(x) = P(y))$

BA3 $(x = y) = \forall z(x(z) = y(z))$

BA4 $(\lambda x.\alpha)(\beta) = \alpha[x \mapsto \beta]$ [β free for x in α]

BA5 $\imath(\mathbf{eqInd}(y)) = y$

The first axiom states that there are only two truth values. The second, that if two objects are equal, then the result of applying a boolean function to both of them is the same. The third, that the identity condition for functions is extensional. **BA4** is β-reduction (both η- and α-reduction are a consequence of the axioms). The fifth axiom states that the property of being equal to an object (**eqInd**) denotes a single object, which is that object.

In addition, in Compositional DRT states and discourse entities must satisfy the following three axioms:

AX1 $\forall i \forall v_\alpha \forall x_\alpha ST_\alpha(v) \rightarrow \exists j(i[v]j \wedge V_\alpha(v,j) = x)$ for $\alpha \in \Theta$

AX2 $\mathrm{ST}_\alpha(\mathrm{v}_\alpha)$ for each discourse entity v_α of type π_α, for $\alpha \in \Theta$;

AX3 $u_y \neq u'_y$ for each two different discourse entities of type π_y, where y = e, τ or ϵ.

Time intervals are ordered by a relation of complete temporal precedence $<$. Four further relations –temporal inclusion $\subseteq$, temporal overlap O, 'immediate precedence' $\trianglelefteq$, and 'partial precedence' $<<$–are defined as follows:

$t_1 \subseteq t_2$ abbreviates $\forall t(t_2 < t \rightarrow t_1 < t) \wedge \forall t(t < t_2 \rightarrow t < t_1)$
$t_1 O t_2$ abbreviates $\exists t(t \subseteq t_1 \wedge t \subseteq t_2)$
$t_1 << t_2$ abbreviates $\forall t(t_2 < t \rightarrow t_1 < t)$
$t_1 \trianglelefteq t_2$ abbreviates $t_1 < t_2 \wedge \neg \exists t_3 t_1 < t_3 < t_2$

The following five axioms on time intervals state that temporal precedence is a strict partial order; that given any two period either they overlap or one of the two precedes the other; and that temporal inclusion $\subseteq$ is anti-symmetric.

AX4 $\forall t \neg t < t$

AX5 $\forall t_1 t_2 t_3((t_1 < t_2 \wedge t_2 < t_3) \rightarrow t_1 < t_3$

AX6 $\forall t_1 t_2 t_3(t_1 < t_2 \vee t_1 O t_2 \vee t_2 < t_1)$

AX7 $\forall t_1 t_2 t_3((t_1 \subseteq t_2 \wedge t_2 \subseteq t_1) \rightarrow t_1 = t_2$

AX8 $\forall t_1 \exists t_2 t_1 \trianglelefteq t_2 \wedge \forall t_1 \exists t_2 t_2 \trianglelefteq t_1$

Each eventuality *e* is associated with a unique time interval $\vartheta(e)$. Because of this association it is possible to define relations among eventualities that reflect those among events: e.g., it is possible to state that $e_1 < e_2$ iff the two events are part of the same world and $\vartheta(e_1) < \vartheta(e_2)$. In addition, Muskens imposes the following axioms on events:

AX9 $\forall e \exists w e$ **in** w

AX10 $\forall e_1 e_2(e_1 << e_2 \rightarrow \forall w(e_2$ **in** $w \rightarrow e_1$ **in** $w))$

AX11 $\forall t \forall w \exists e e$ **at** $t \wedge e$ **in** w[68]

[68] e **at** t is an abbreviation for $\vartheta(e) = t$.

References

Abney, S. (1991). Parsing by chunks. In R. Berwick, S. Abney, and C.Tenny, editors, *Principle-based Parsing*. Kluwer, Dordrecht.

Alshawi, H. (1990). Resolving quasi-logical forms. *Computational Linguistics*, **16**(3), 133–144.

Alshawi, H., editor (1992). *The Core Language Engine*. The MIT Press.

Alshawi, H. and Crouch, R. (1992). Monotonic semantic interpretation. In *Proc. 30th. ACL*, pages 32–39, University of Delaware.

Andrews, P. B. (1986). *An Introduction to Mathematical Logic and Type Theory: to Truth through Proof*. Academic Press, New York.

Asher, N. (1993). *Reference to Abstract Objects in English*. D. Reidel, Dordrecht.

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

Bacchus, F. (1990). *Representing and Reasoning with Probabilistic Knowledge*. The MIT Press, Cambridge, MA.

Bacchus, F., Grove, A. J., Halpern, J. Y., and Koller, D. (1993). Statistical foundations for default reasoning. In *Proc. of the 13th IJCAI*, pages 563–569, Chambery.

Backofen, R., Rogers, J., and Vijay-Shanker, K. (1995). A first-order axiomatization of the theory of finite trees. *Journal of Logic, Language and Computation*.

Bard, E. G., Robertson, D., and Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language*, **72**(1), 32–68.

Barwise, J. (1987). Noun phrases, generalized quantifiers and anaphora. In Gärdenfors (1987), pages 1–30.

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

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

Bos, J., Mastenbroek, E., McGlashan, S., Millies, S., and Pinkal, M. (1994). The Verbmobil semantic formalism (version 1.3). Verbmobil-report 6, Universität des Saarlandes, Saarbruecken.

Brennan, S., Friedman, M., and Pollard, C. (1987). A centering approach to pronouns. In *Proc. of the 25th ACL*, pages 155–162.

Brewka, G. (1991). *Nonmonotonic Reasoning: Logical Foundations of Commonsense*. Cambridge University Press, Cambridge.

Bronneberg, W., Bunt, H. C., Landsbergen, S. P. J., Scha, R., Schoenmakers, W., and Utteren, E. v. (1979). The question-answering system PHLIQA1. In L. Bolc, editor, *Natural Communication with Computers, vol. 2*. Carl Hanswer Verlag, Munich and Vienna.

Carpenter, B. (1992). *The Logic of Typed Feature Structures*. Cambridge.

Carpenter, B. (1994). Quantification and scoping: a deductive account. In *Proc. of the 13th West Coast Conference on Formal Linguistics*.

Carpenter, B. (1998). *Type-Logical Semantics*. MIT Press.

Chater, N., Pickering, M. J., and Milward, D. (1995). What is incremental interpretation? In Milward, D. R. and Sturt, P. (editors), Incremental Interpretation, Edinburgh Working Papers in Cognitive Science 11, University of Edinburgh, Centre for Cognitive Science.

Clark, H. H. (1977). Bridging. In P. N. Johnson-Laird and P. Wason, editors, *Thinking: Readings in Cognitive Science*. Cambridge University Press, London and New York.

Cooper, R. (1975). *Montague's Semantic Theory and Transformational Syntax*. Ph.D. thesis, University of Massachusetts, Amherst.

Cooper, R. (1983). *Quantification and Syntactic Theory*. D. Reidel Publishing Company, Dordrecht, Holland.

Copestake, A. and Briscoe, T. (1995). Semi-productive polysemy and sense extension. *Journal of Semantics*, **12**(1), 15–68. Special Issue on Lexical Semantics.

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

Corley, S. (1998). *A Statistical Model of Human Lexical Category Disambiguation*. Ph.D. thesis, University of Edinburgh, Cognitive Science.

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, L. Karttunen, and A. M. Zwicky, editors, *Natural Language Parsing: Psychological, Computational and Theoretical perspectives*, pages 320–358. Cambridge University Press, New York.

Cruse, D. A. (1995). Polysemy and related phenomena from a cognitive linguistic viewpoint. In P. Saint-Dizier and E. Vegas, editors, *Computational Lexical Semantics*, chapter 2, pages 33–49. Cambridge University Press.

Cutler, A. and Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 113–121.

Davidson, D. (1967). The logical form of action sentences. In N. Rescher, editor, *The Logic of Decision and Action*, pages 81–95. University of Pittsburgh Press, Pittsburgh.

Davies, B. (1998). *An Empirical Examination of Cooperation, Effort and Risk in Task-Oriented Dialogue*. Ph.D. thesis, University of Edinburgh, Linguistics.

Dowty, D. R., Wall, R. E., and Peters, S. (1981). *Introduction to Montague Semantics*. D. Reidel, Dordrecht, Holland.

Duffy, S., Morris, R., and Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, **27**, 429–446.

Evans, G. (1980). Pronouns. *Linguistic Inquiry*, **11**(2), 337–362.

Fillmore, C. J. (1988). The mechanisms of "construction grammar". In S. Axmaker, A. Jaisser, and H. Singmaster, editors, *Proceedings of the 14th Annual Meeting of the Berkeley Linguistics Society*, pages 35–55, Berkeley, CA.

Forster, K. I. and Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, **12**, 627–635.

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

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

Frazier, L. and Clifton, C. (1997). Construal: overview, motivation, and some new evidence. *Journal of Psycholinguistic Research*, **26**(3), 277–295.

Frazier, L. and Rayner, K. (1987). Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language*, **26**, 505–526.

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

Gärdenfors, P., editor (1987). *Generalized Quantifiers*. D. Reidel, Dordrecht, The Netherlands.

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

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

Garrod, S. C., Freudenthal, D., and Boyle, E. (1993). The role of different types of anaphor in the on-line resolution of sentences in a discourse. *Journal of Memory and Language*, **32**, 1–30.

Geffner, H. and Pearl, J. (1990). A framework for reasoning with defaults. In H. E. K. Jr., R. Loui, and G. Carlson, editors, *Knowledge Representation and Defeasible Reasoning*. Kluwer, Dordrecht.

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*, **32**, 99–156.

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

Gordon, P. C. (1993). Computational and psychological models of discourse. In H. H. Brownell and Y. Joanette, editors, *Narrative discourse in neurologically impaired and normal aging adults*, chapter 2, pages 23–46. Singular Publishing Group, San Diego.

Gordon, P. C., Grosz, B. J., and Gillion, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, **17**, 311–348.

Gorrell, P. (1995). *Syntax and Parsing*. Cambridge University Press.

Grice, H. P. (1957). Meaning. *Philosophical Review*, **66**, 377–388. also in D.A. Steinberg and L. A. Jakobovits (eds.), *Semantics*, Cambridge University Press, 1971.

Groenendijk, J. and Stokhof, M. (1991). Dynamic Predicate Logic. *Linguistics and Philosophy*, **14**, 39–100.

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

Grosz, B., Joshi, A., and Weinstein, S. (1983). Providing a unified account of definite noun phrases in discourse. In *Proc. ACL-83*, pages 44–50.

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

Haddock, N. (1988). *Incremental Semantics and Interactive Syntactic Processing*. Ph.D. thesis, Dept. of AI and Centre for Cognitive Science, University of Edinburgh.

Heim, I. (1982). *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. thesis, University of Massachusetts at Amherst.

Henkin, L. (1950). Completeness in the theory of types. *Journal of Symbolic Logic*, **15**, 81–91.

Hindle, D. and Rooth, M. (1993). Structural ambiguity and lexical relations. *Computational Linguistics*, **19**(1), 103–120. Special Issue on Using Large Corpora, I.

Hirst, G. (1987). *Semantic Interpretation and the Resolution of Ambiguity*. Studies in Natural Language Processing. Cambridge University Press, Cambridge, UK.

Hobbs, J. R. (1982). Representing ambiguity. In *Proc. WCCFL*, pages 15–28.

Hobbs, J. R. (1983). An improper treatment of quantification in ordinary English. In *Proc. ACL-83*, pages 57–63, Cambridge, MA.

Hobbs, J. R. (1985). Granularity. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, pages 432–435, Los Angeles, California. International Joint Committee on Artificial Intelligence.

Hobbs, J. R. and Martin, P. (1987). Local pragmatics. In *Proc. IJCAI-97*, pages 520–523, Milano, Italy.

Hobbs, J. R. and Shieber, S. M. (1987). An algorithm for generating quantifier scopings. *Computational Linguistics*, **13**(1-2), 47–63.

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

Hurum, S. (1988). Handling scope ambiguities using domain-independent heuristics. Technical Report TR 88-12, University of Alberta, Edmonton, Canada.

Hwang, C. H. and Schubert, L. K. (1992). Tense trees as the “fine structure” of discourse. In *Proc. ACL-92*, pages 232–240, Newark, DE.

Hwang, C. H. and Schubert, L. K. (1993). Episodic logic: A situational logic for natural language processing. In P. Aczel, D. Israel, Y. Katagiri, and S. Peters, editors, *Situation Theory and its Applications, v.3*, pages 303–338. CSLI.

Ioup, G. (1975). Some universals for quantifier scope. In J. Kimball, editor, *Syntax and Semantics 4*, pages 37–58. Academic Press, New York.

Johnson, M. (1988). *Attribute-value logic and the theory of grammar*. Lecture Notes. CSLI, Stanford, CA.

Johnson, M. K., Bransford, J. D., and Solomon, S. (1973). Memory for tacit implications of sentences. *Journal of Experimental Psychology*, **98**, 203–205.

Joshi, A. (1987). Introduction to tree adjoining grammars. In A. Manaster-Ramis, editor, *Mathematics of Language*. John Benjamins, Amsterdam.

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

Kálmán, L. (1990). Deferred information: The semantics of commitment. In L. Kálmán and L. Pólos, editors, *Papers from the Second Symposium on Logic and Language*, pages 125–148. Akadémiái Kiadó, Budapest.

Kameyama, M. (1998). Intra-sentential centering: A case study. In M. A. Walker, A. K. Joshi, and E. F. Prince, editors, *Centering Theory in Discourse*, chapter 6, pages 89–112. Oxford.

Kameyama, M., Passonneau, R., and Poesio, M. (1993). Temporal centering. In *Proc. of the 31st ACL*, pages 70–77, Columbus, OH.

Kamp, H. (1990). Prolegomena to a structural account of belief and other attitudes. In C. A. Anderson and J. Owens, editors, *Propositional Attitudes—The Role of Content in Logic, Language, and Mind*, chapter 2, pages 27–90. University of Chicago Press and CSLI, Stanford.

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

Keller, F. (1996). How do humans deal with ungrammatical input? In D. Gibbon, editor, *Natural Language Processing and Speech Technology*, pages 27–34. Mouton de Gruyter, Berlin.

Kempson, R. and Cormack, A. (1981). Ambiguity and quantification. *Linguistics and Philosophy*, **4**(2), 259–310.

König, E. (1989). Parsing as natural deduction. In *Proc. of the 27th ACL*, pages 272–279, Vancouver.

Kratzer, A. (1977). What 'must' and 'can' must and can mean. *Linguistics and Philosophy*, **1**, 337–355.

Kripke, S. A. (1977). Speaker reference and semantic reference. In P. A. French, T. E. Uehling, and H. K. Wettstein, editors, *Contemporary Perspectives in the Philosophy of Language*, pages 6–27. University of Minnesota Press, Minneapolis.

Kurtzman, H. S. and MacDonald, M. C. (1993). Resolution of quantifier scope ambiguities. *Cognition*, **48**, 243–279.

Lakoff, G. P. (1971). Semantic interpretation in generative grammar. In D. A. Steinberg and L. A. Jakobovits, editors, *Semantics: An interdisciplinary reader in philosophy, linguistics, anthropology, and psychology*. Cambridge University Press, Cambridge, UK. ISBN (paper) 0 521 09881 5.

Lambek, J. (1958). The mathematics of sentence structure. *American Mathematical Monthly*, **65**, 154–170. Reprinted in W. Buszkowski, W. Marciszewski and J. van Benthem (eds.), (1989), *Categorial Grammar*. Amsterdam: Benjamins.

Lascarides, A. and Asher, N. (1993). Temporal interpretation, discourse relations and commonsense entailment. *Linguistics and Philosophy*, **16**(5), 437–493.

Lascarides, A., Briscoe, E. J., Asher, N., and Copestake, A. (1996a). Persistent default unification. *Linguistics and Philosophy*, **19**(1), 1–89.

Lascarides, A., Copestake, A., and Briscoe, E. J. (1996b). Ambiguity and coherence. *Journal of Semantics*, **13**(1), 41–66.

Levinson, S. (1983). *Pragmatics*. Cambridge University Press.

Lewin, I. (1990). A quantifier scoping algorithm without a free variable constraint. In *Proc. COLING*, pages 190–194.

Lewis, D. K. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, **8**, 339–359.

Luce, P. A. (1986). *Neighborhoods of words in the mental lexicon*. Ph.D. thesis, Indiana University, Bloomington.

Lyons, J. (1995). *Linguistic Semantics*. Cambridge University Press.

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

Marcus, M. (1980). *A Theory of Syntactic Recognition for Natural Language*. The MIT Press, Cambridge, MA.

Marcus, M., Hindle, D., and Fleck, M. (1983). D-theory: Talking about talking about trees. In *Proc. of the 21st Annual Meeting of the ACL*.

Marslen-Wilson, W. (1984). Function and process in spoken-word recognition. In E. Bouma and P. G. Bouwhuis, editors, *Attention and Performance: Control of Language Processes*, pages 125–150. Lawrence Erlbaum, Hillsdale, NJ.

Meyer, D. E. and Schvaneveldt, R. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, **90**, 227–235.

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

Milward, D. and Cooper, R. (1994). Incremental interpretation: Applications and relationship to dynamic semantics. In *Proc. COLING-94*, Kyoto.

Milward, D. and Sturt, P. (1995). Incremental interpretation. Edinburgh Working Papers in Cognitive Science 11, University of Edinburgh, Centre for Cognitive Science.

Montague, R. (1973). The proper treatment of quantification in English. In K. e. a. Hintikka, editor, *Approaches to Natural Language*, pages 221–242. D. Reidel, Dordrecht. Reprinted in Thomason (1974).

Moore, R. C. (1981). Problems in logical form. In *Proc. of the 19th ACL*, pages 107–124, Stanford, CA.

Moore, R. C. (1989). Unification-based semantic interpretation. In *Proc. of the 27th ACL*, pages 33–41.

Moran, D. (1988). Quantifier scoping in the SRI Core Language Engine. In *Proc. ACL-88*, pages 33–40, Buffalo, NY.

Muskens, R. (1994). A compositional discourse representation theory. In P. Dekker and M. Stokhof, editors, *Proceedings of the 9th Amsterdam Colloquium*, pages 467–486.

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

Muskens, R. (1995b). Tense and the logic of change. In U. Egli, P. Pause, C. Schwarze, A. von Stechow, and G. Wienold, editors, *Lexical Knowledge in the Organization of Language*, pages 147–183. John Benjamins, Amsterdam / Philadelphia.

Neale, S. (1990). *Descriptions*. The MIT Press, Cambridge, MA.

Nunberg, G. D. (1978). *The pragmatics of reference*. Indiana University Linguistics Club.

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

Oaksford, M. and Chater, N. (1991). Against logicist cognitive science. *Mind and Language*, **1**(1), 1–38.

Partee, B. H. (1972). Opacity, coreference, and pronouns. In D. Davidson and G. Harman, editors, *Semantics for Natural Language*, pages 415–441. D. Reidel, Dordrecht, Holland.

Partee, B. H. and Rooth, M. (1983). Generalized conjunction and type ambiguity. In R. Bauerle, C. Schwarze, and A. von Stechow, editors, *Meaning, Use and Interpretation of Language*. Walter de Gruyter, Berlin, West Germany.

Paul, S. T., Kellas, G., Martin, M., and Clark, M. B. (1992). The influence of contextual features on the activation of ambiguous word meanings. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **18**, 703–717.

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

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

Pereira, F. C. N. and Pollack, M. E. (1991). Incremental interpretation. *Artificial Intelligence*, **50**, 37–82.

Pereira, F. C. N. and Shieber, S. M. (1987). *Prolog and Natural-Language Analysis*. CSLI Lecture Notes. CSLI, Stanford.

Pinkal, M. (1995). *Logic and Lexicon*. D. Reidel, Dordrecht.

Poesio, M. (1991). Relational semantics and scope ambiguity. In J. Barwise, J. M. Gawron, G. Plotkin, and S. Tutiya, editors, *Situation Semantics and its Applications, vol.2*, chapter 20, pages 469–497. CSLI, Stanford, CA.

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

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

Poesio, M. (1996b). Disambiguation as (defeasible) reasoning about underspecified representations. In P. Dekker and M.Stokhof, editors, *Proc. of the Tenth Amsterdam Colloquium*, pages 607–625. ILLC.

Poesio, M. (1996c). Ambiguity and underspecification. In M. Bierwisch, R. Blutner, and R. van der Sandt, editors, *From underspecification to interpretation: Proceedings of the Berlin Workshop on Underspecification*. IBM Deutschland: Working Papers of the Institute for Logic and Linguistics, Heidelberg.

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

Poesio, M. and Reyle, U. (1999). Partially resolved interpretations. In *Proc. of the Paris Conference on Syntax and Semantics*.

Poesio, M. and Stevenson, R. (To appear). *Salience: Computational Models and Psychological Evidence*. Cambridge University Press, Cambridge and New York.

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

Poesio, M. and Vieira, R. (1998). A corpus-based investigation of definite description use. *Computational Linguistics*, **24**(2), 183–216. Also available as Research Paper CCS-RP-71, Centre for Cognitive Science.

Pollard, C. and Sag, I. A. (1987). *Information-Based Syntax and Semantics*. Number 13 in Lecture Notes. CSLI, Stanford, CA.

Pollard, C. and Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*. University of Chicago Press, Chicago, CA.

Poole, D. (1993). Probabilistic Horn abduction and Bayesian networks. *Artificial Intelligence*, **64**(1), 81–129.

Prince, A. and Smolensky, P. (1993). Optimality theory: Constraint interaction in generative grammar. Technical Report 2, Rutgers University, Centre for Cognitive Science.

Pulman, S. (1994). A computational theory of context dependence. In *Proc. of the 1st International Workshop on Computational Semantics*, pages 161–170. University of Tilburg.

Pustejovsky, J. (1991). The generative lexicon. *Computational Linguistics*, **17**(4), 409–441.

Raskin, V. (1985). *Semantic Mechanisms of Humor*. D. Reidel, Dordrecht and Boston.

Rayner, K. and Morris, R. K. (1991). Comprehension processes in reading ambiguous sentences: Reflections from eye movements. In G. B. Simpson, editor, *Understanding word and sentence*, pages 175–198. North-Holland.

Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence*, **13**(1–2), 81–132.

Reiter, R. and Criscuolo, G. (1981). On interacting defaults. In *Proc. 7th IJCAI*, pages 270–276.

Reuland, E. J. and ter Meulen, A. G. B., editors (1987). *The Representation of (In)definiteness*. The MIT Press, Cambridge, MA.

Reyle, U. (1993). Dealing with ambiguities by underspecification: Construction, representation and deduction. *Journal of Semantics*, **10**, 123–179.

Reyle, U. (1996). Co-indexing labeled DRSs to represent and reason with ambiguities. In K. van Deemter and S. Peters, editors, *Semantic Ambiguity and Underspecification*, chapter 10, pages 239–268. CSLI, Stanford.

Richards, I. A. (1936). *The Philosophy of Rhetoric*. Oxford University Press, London.

Rooth, M. (1985). *Association with Focus*. Ph.D. thesis, University of Massachusetts, Amherst.

Rooth, M. (1987). Noun Phrase Interpretation in Montague Grammar, File Change Semantics, and Situation Semantics. In Gärdenfors (1987), pages 237–268.

Rosch, E. (1975). Cognitive representation of semantic categories. *Journal of Experimental Psychology: General*, **104**, 192–233.

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

Schubert, L. K. and Pelletier, F. J. (1982). From English to Logic: Context-free computation of 'conventional' logical translations. *American Journal of Computational Linguistics*, **10**, 165–176.

Schütze, H. (1997). *Ambiguity Resolution in Language Learning*. CSLI, Stanford.

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

Selman, B. and Kautz, H. (1990). Model-preference default theories. *Artificial Intelligence*, **45**, 287–322.

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

Shieber, S., Schabes, Y., and Pereira, F. (1995). Principles and implementation of deductive parsing. *Journal of Logic Programming*, **24**(1-2), 3–36.

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

Simpson, G. B. (1994). Context and the processing of ambiguous words. In M. A. Gernsbacher, editor, *Handbook of Psycholinguistics*, chapter 10, pages 359–374. Academic Press.

Smyth, R. (1994). Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research*, **23**, 197–229.

Stalnaker, R. (1973). Presupposition. *Journal of Philosophical Logic*, **2**, 447–457.

Stevenson, R. J. and Vitkovitch, M. (1986). The comprehension of anaphoric relations. *Language and Speech*, **29**, 335–360.

Stevenson, R. J., Nelson, A. W. R., and Stenning, K. (1993). Strategies in pronoun comprehension. In *Proceedings of the 15th Annual Meeting of the Cognitive Science Society*, pages 976–981.

Stevenson, S. (1994). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research*, **23**(4), 295–321.

Stolcke, A. (1994). *Bayesian Learning of Probabilistic Language Models*. Ph.D. thesis, Computer Science Division, University of California, Berkeley.

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

Su, S. P. (1994). *Lexical Ambiguity in Poetry*. Longman, London.

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

Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential context. *Journal of Memory and Language*, **27**, 324–340.

Tanenhaus, M. K., Leiman, J. M., and Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior*, **18**(4), 427–440.

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

Thomason, R. H., editor (1974). *Formal Philosophy: Selected Papers of Richard Montague*. Yale University Press, New York.

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

Trueswell, J. C., Tanenhaus, M., and Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **19**(3), 528–553.

van Benthem, J. F. A. K. (1982). Five easy pieces. In A. ter Meulen, editor, *Studies in Model Theoretic Semantics*. Foris, Dordrecht.

van Deemter, K. (1991). *On the Composition of Meaning*. Ph.D. thesis, University of Amsterdam.

van Deemter, K. (1996). Towards a logic of ambiguous expressions. In K. van Deemter and S. Peters, editors, *Semantic Ambiguity and Underspecification*, chapter 9, pages 203–237. CSLI, Stanford.

van Deemter, K. (1998). Ambiguity and idiosyncratic interpretation. *Journal of Semantics*, **15**(1), 5–36.

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

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

Verkuyl, H. J. and van der Does, J. (1991). The semantics of plural noun phrases. IILC paper.

von Stechow, A. and Wunderlich, D., editors (1991). *Semantik : ein internationales Handbuch der zeitgenossischen Forschung*. Walter de Gruyter, Berlin.

Weinberg, A. (1993). Parameters in the theory of sentence processing: Minimal Commitment Theory goes East. *Journal of Psycholinguistic Research*, **22**(3), 339–364.

Woods, W. A., Kaplan, R., and Nash-Webber, B. (1972). The lunar sciences natural language information system: Final report. Report 2378, BBN, Cambridge, MA.

Zwicky, A. and Sadock, J. (1975). Ambiguity tests and how to fail them. In J. Kimball, editor, *Syntax and Semantics 4*, pages 1–36. Academic Press, New York.

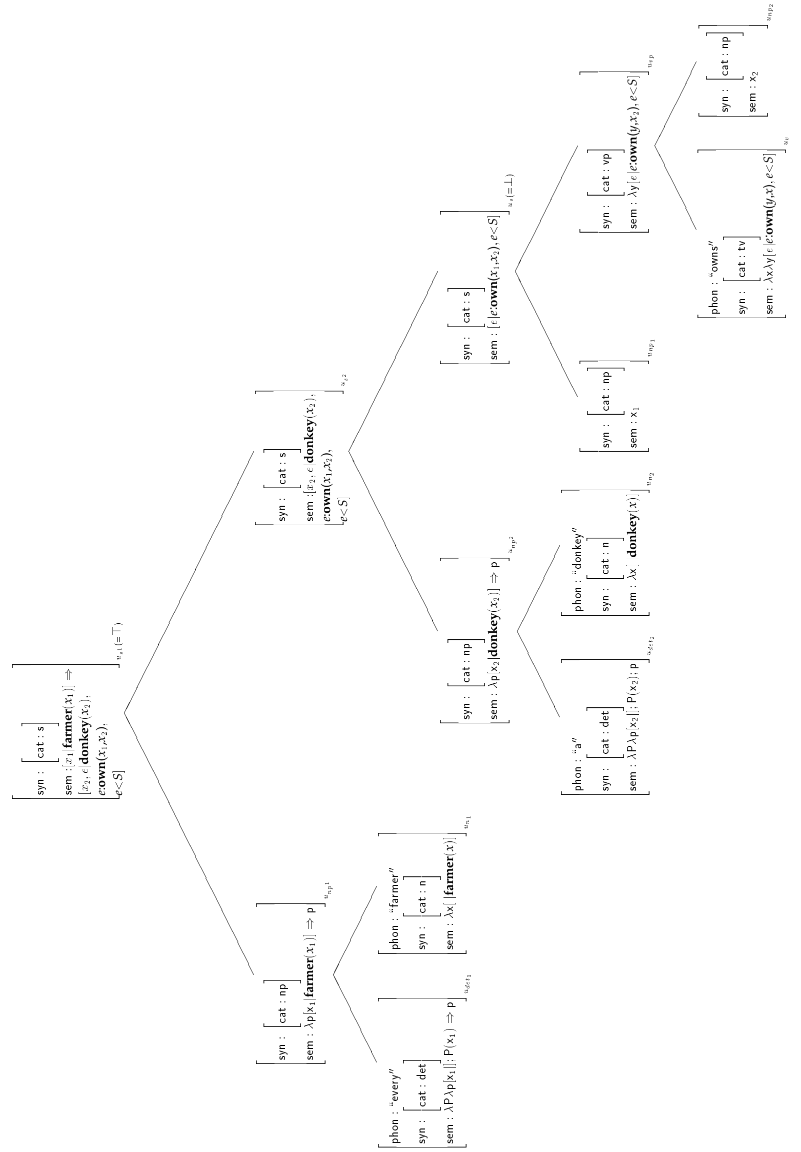


Figure 4: The structural relations among utterances inferred while deriving one interpretation of *Every farmer owns a donkey*.

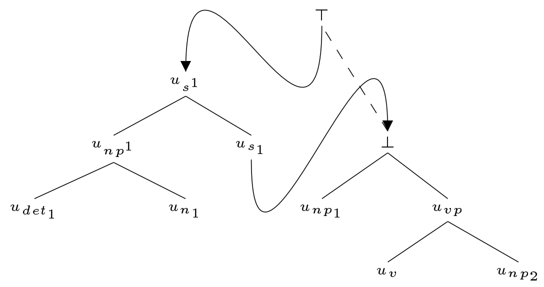


Figure 5: Adjunction of the 'stored' component of *every farmer*.