

Definite Descriptions, Focus Shift, and a Theory of Discourse Interpretation

Massimo Poesio
Centre for Cognitive Science
University of Edinburgh
poesio@cogsci.ed.ac.uk

Abstract

The model of definite description interpretation in [Poesio, 1993] is revised to avoid the previous theory's assumption that the interpretation of definites occurs after scope interpretation has taken place. The processes involved in the interpretation of a definite are formalized as DRS construction rules applying in parallel to generate alternative hypotheses out of underspecified representations. Interpreting a definite requires identifying the particular situation, or *discourse topic*, that includes an object of that type. 'Focus' effects are formulated in terms of situations and their hierarchical organization.

1 Ingredients of a Theory of Definite Description Interpretation

A theory of definite description (DD) interpretation must build on a general theory of discourse interpretation, i.e., a theory of how hypotheses about the intended meaning of an utterance are arrived at, and must include:

1. A theory of the semantics of definite descriptions, i.e., of what DDs mean, and what part of their interpretation has to be recovered in context; and
2. A theory of the pragmatics of definite descriptions, i.e., how the contextually dependent aspects of the meaning of DDs are recovered in a discourse situation. An aspect of the discourse situation that has been observed to play an especially important role in the interpretation of DDs is the *attentional state* of the conversational participants [Grosz and Sidner, 1986], i.e., on what subset of their shared knowledge they are focusing on.

In [Poesio, 1993] I address these last two points in some detail, and present a model of definite description interpretation and the attentional state that includes (i) a theory of the contextually dependent aspects of definite description interpretation, based on Hawkins' *location theory of definite descriptions* [1978] and formalized in situation-theoretic terms using Episodic Logic [Hwang and Schubert, 1993]; (ii) a revision of Grosz and Sidner's theory of discourse structure, reformulated using situations to model the semi-formal notions of 'focus spaces' and 'discourse segments'; and (iii) a formal description of the discourse processes involved in definite description interpretation, according to which both the principles for resolving the contextually dependent aspects of definite descriptions, and those used to inferring the current focus of attention, are formulated as default inference rules in the sense of Reiter [1980].

The weaker aspect of the proposal in [Poesio, 1993] is the reliance on an ad-hoc theory of discourse interpretation. A particularly unsatisfactory aspect of the theory developed in that paper is the assumption (motivated by reasons of simplicity) that definite description interpretation takes place after every other aspect of the interpretation of a sentence, including the scope of operators, had been resolved. This aspect of the proposal was unsatisfactory for two reasons: first of all, it is well-known from psychological experiments that the interpretation of definite descriptions takes place quite early, in fact may take place even before parsing is completed, as demonstrated by the impact of definite interpretation on garden-path sentences [Crain and Steedman, 1985]. Secondly, the choice of an interpretation for definite descriptions affects their scope, but not viceversa. Consider the following example from Heim [1982]:

- (1) If a cat sees a dog, the cat meows.

the scope of the definite *the cat* in the consequent of the conditional in (1) is clearly affected by the choice of its antecedent. If the antecedent is taken to be some cat in the context, the definite takes wide scope over the (implicit) adverb of quantification ('every time that ... ') introduced by the conditional, and the sentence is interpreted as 'there is cat that meows every time a(nother) cat sees a dog; the preferred interpretation, according to which the cat that meows is the same cat that saw a dog, is obtained if *the cat* is interpreted as anaphoric to *a cat* in the antecedent of the conditional.

In the revised model of definite description interpretation I have been developing after the paper just mentioned, and that I sketchily describe here, these problems with the theory presented in [Poesio, 1993] are resolved by reformulating discourse interpretation as a reasoning process over *underspecified representations*, representations that encode sets of possible interpretations of a sentence. Several unclear aspects of the earlier model have also been clarified, and my early assumptions concerning the semantics of definite descriptions have been revised. In this paper, I first introduce a psychologically more plausible model of discourse interpretation called *Conversation Representation Theory*; I then discuss my hypothesis about the semantics and pragmatics of definite descriptions, and finally I briefly describe how the model is applied to formulate an account of the interpretation of 'visible situation' definite descriptions. The discourse interpretation system that builds upon the theory presented here, called SAD-93, is described in [Poesio,

1994a].

2 Conversation Representation Theory

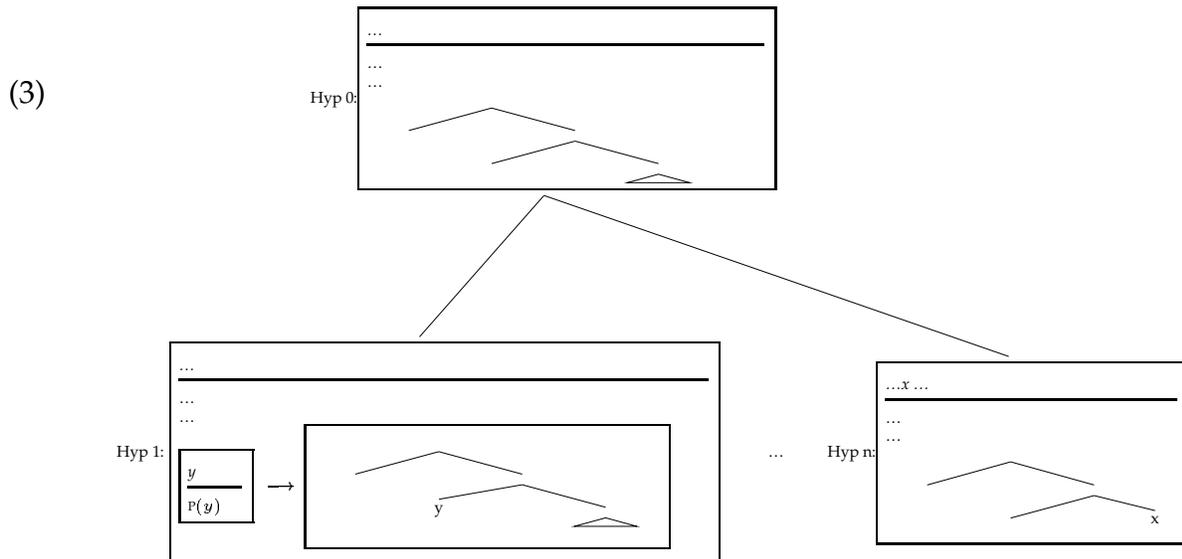
Conversation Representation Theory (CRT) is a theory of discourse representation and interpretation developed in [Poesio, 1994a] to serve as the foundation of an account of discourse interpretation and scope disambiguation.

I concentrate here on the aspects of CRT relevant for the task at hand. One of these is the assumption that discourse interpretation is a process during which several hypotheses are generated in parallel, and then filtered on the basis of contextual information. There is evidence suggesting that both lexical processing and syntactic processing take place in this fashion [Seidenberg *et al.*, 1982; Swinney, 1979; Kurtzman, 1985; Crain and Steedman, 1985; Gibson, 1991]; Kurtzman and MacDonald [1993] suggest a similar model for scope disambiguation as well. As far as reference interpretation is concerned, there is some evidence that all pragmatically available referents become active before a referent is identified (see, e.g., [Spivey-Knowlton *et al.*, 1994]). The parallel processing hypothesis is also consistent with the well-known fact that a sentence can deliberately be made ambiguous for humorous or rhetorical purposes, as in the following example (quoted in [Raskin, 1985], p. 25-26):

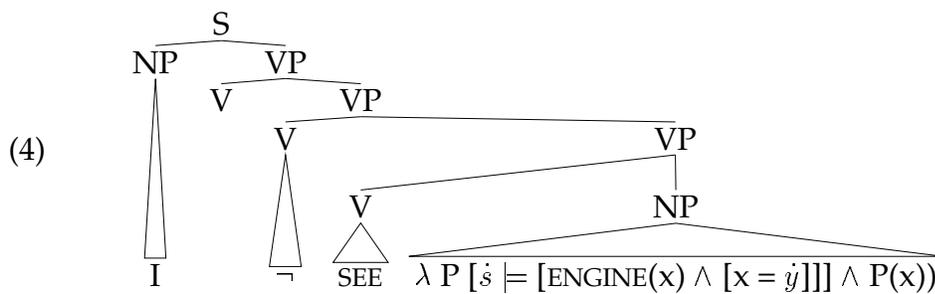
(2) The first thing that strikes a stranger in New York is a big car.

The joke relies on two assumptions about human processing: that the clause *the first thing that strikes a stranger in New York* gets interpreted before the end of the sentence, *strikes* receiving the 'surprise' interpretation; secondly, that the reader is able to go back, arrive at a second interpretation, and entertain them both simultaneously.

Another assumption shared by much research on discourse processing is that interpretation is the process of integrating an utterance's interpretation into the existing context (e.g., [Garrod, 1993]). Together, the two assumptions of parallel processing and of interpretation as context modification lead to a model of interpretation that can be summarized as in the diagram in (3).



The initial hypothesis, as well as the hypotheses obtained about discourse interpretation, are hypotheses about the interpretation of a sentence given a certain context, represented in the form of a ‘box’ in the diagram. The input to discourse interpretation is an *underspecified interpretation*, i.e., an interpretation that ‘packs’ together several alternative interpretations. Here as in [Poesio, 1994a], I ignore attachment ambiguities and lexical ambiguities to simplify matters, and assume that the input to discourse interpretation consists of syntactic information together with the semantic translation of lexical items. This information can be represented as a syntactic tree whose leaves have been decorated with their semantic translation. Assuming an interpretation for lexical items of the form proposed in Montague Grammar, the sentence *I don’t see the engine* can be represented as in (4). (The translation of definites is discussed more in detail below.)



The diagram in (3) also illustrates the fact that each interpretive step consists of the generation of alternative hypotheses, each of which is again an hypothesis ‘relative to a context’. An hypothesis about a context can be represented as in Discourse Representation Theory (DRT) [Kamp and Reyle, 1993], by means of *Discourse Representation Structures* (DRSs). A DRS is a pair $\langle U, C \rangle$, where U is a set of discourse referents, and C is a set of conditions on (predications about) these referents.

In [Kamp and Reyle, 1993], the procedure used to assign truth conditions to a sentence is specified in terms of *construction rules*, operations from DRSs to DRSs. I propose in [Poesio, 1994a] that the format adopted by Kamp and Reyle to specify their semantic

interpretation procedure can be recycled to formalize discourse interpretation, in the sense that the operations that result in new hypotheses can be formalized as construction rules. A construction rule is, essentially, a generalization of the traditional notion of inference rule, in that it can augment both the set of conditions and the set of discourse referents of a DRS (whereas a normal inference rule only augments a set of conditions). More precisely, construction rules as I propose to use them are generalizations of Reiter's *defaults* [Reiter, 1980]: the result of the application of a construction rule is a new *extension* of the theory consisting of the initial hypothesis together with the default inference rules, and different extensions need not be mutually consistent. Each extension is more complete (less ambiguous, in the case of scope disambiguation) than the initial hypothesis. Some of these extensions may be inconsistent, and other extensions may be equivalent, at least for the current purposes (see below); otherwise, a conflict is obtained, that may be resolved in different ways.

The underspecified representation in (4) denotes the set of propositions that can be obtained by (i) assigning a scope to all the operators—quantifiers, negation, etc.—by means, say, of a procedure like Cooper's Store [Cooper, 1983]; and (ii) assigning a value to all *parameters*, of which (4) contains two, \dot{y} and \dot{s} . (The translation of the definite *the engine* in (4) is explained below.) Parameters are used to specify the aspects of a sentence's meaning that need to be resolved in context by discourse interpretation. Semantically, a parameter denotes a set of functions from situations to the objects made 'available' in the discourse situation. For example, if the objects $a_1 \dots a_n$ are 'made available' in the sense described below in the discourse situation d , a parameter like \dot{y} in (4) denotes in d a set $\{f_1, \dots, f_n\}$ of functions from situations to $a_1 \dots a_n$.

A parameter is *anchored* if only one among the functions in its denotation results in a consistent interpretation of the DRS in which the parameter occurs. A parameter can be anchored by means of equality statements of the form $[\dot{x} = a]$, where a is not parametric, or is already anchored: such equality statements make all but one of the interpretations of the parameter inadmissible. Once a parameter is anchored, it can be 'replaced' by a term that denotes the one function among those in the interpretation of the parameter that does not result in an inconsistent interpretation. The goal of discourse interpretation is to anchor parameters. A conversation is infelicitous unless all parameters can be anchored: the referents of all pronouns and definite descriptions have been identified the domain of quantification of all quantifiers has been appropriately restricted, and so forth: so much so that listeners appear to be ready to *accommodate* new information (e.g., introduce into the discourse some otherwise unspecified antecedent for a pronoun) rather than leave the interpretation parametric. This intuition, first noted by Lewis [1979], can be formalized by the following condition:

Condition on Discourse Interpretation A discourse represented by the root DRS K is infelicitous unless all parameters occurring in K are anchored by the end of discourse interpretation.

The set of hypotheses that result from discourse interpretation and satisfy the Condition on Discourse Interpretation is filtered and organized by plausibility on the basis of

commonsense knowledge. The preferred interpretation of (5), for example, is the one where the pronoun *it* refers to the serial port, because these are usually found in the back of computers.

(5) Hook up the cable to the serial port. *It* is on the back of the computer.

Commonsense reasoning may also tell us that some extensions are equivalent for the purposes at hand, thus can be merged. An example from the TRAINS corpus¹ is shown in (6): even though *it* can refer either to *the engine* or to *the boxcar*, and therefore two extensions could be obtained by discourse interpretation, the difference between these two extensions would be immaterial as far as the plan is concerned, because moving one object would necessarily entail moving the other; the two hypotheses can therefore be merged.

(6) Hook up *the engine* to *the boxcar*, and move *it* to Avon.

An ambiguity is *perceived* when commonsense knowledge cannot rank or merge the hypotheses obtained by discourse interpretation. The precise details of how commonsense knowledge is used to ‘filter’ and ‘merge’ hypotheses are still an open question; the commonsense knowledge-controlled ‘plausibility ranking’ among hypotheses can be formally modeled by means of a priority ranking among models similar to that proposed by Shoham [Shoham, 1988].²

3 The Common Ground as the Representation of a Discourse Situation: Discourse Topics

It is well known from the literature on discourse interpretation [Reichman, 1985; Grosz and Sidner, 1986] that conversations are not ‘about’ a single topic, and that this affects the choice of referents, especially when definite descriptions are concerned. In the TRAINS conversations, for example, that involve two people concerned with the development of a transportation plan, there are at least two topics of conversation: the current state of the world (e.g., which objects are located where), and the plan being developed. A definite description like *the engine* may be interpreted as referring either to an engine in the world, or to an engine mentioned as part of the plan, whose identity need not have been specified. Thus, a sentence such as *The engine is now at Avon* can either be interpreted as stating that an engine is at Avon in the current state of affairs, or as stating that as a consequence of some actions in the plan (yet to be executed), an engine is now located at Avon.

Clearly, part of the task of discourse interpretation is to locate a definite within one of the different topics of conversation. In [Poesio, 1993], I argue that there is a natural way to obtain a model of discourse that reflects the existence of multiple discourse topics. The approach developed there (and adopted here as well) is based on the perspective on context developed by Barwise and Perry [1983], according to whom the participants

to a conversation make use of the knowledge that they are themselves part of a *discourse situation* in which they perform *conversational actions*.

The notion of common ground used in Conversation Representation Theory is not the same as the notion used in DRT. The root DRS is not interpreted in CRT as a characterization of a situation being described by the discourse, but as a characterization of the *discourse* situation itself: in other words, the conditions included in the root DRS describe facts that hold of the discourse situation, as opposed to facts that holds of the situation described by an utterance. These facts include information about the existence of a set of *discourse topics*, that are referred to by the utterances that are included in the discourse situation.

I argue in [Poesio, 1993] that if we assume that discourse topics are situations³, we get an explanation for a number of the observations about the pragmatics of definite descriptions reported in the literature that does not require formal tools other than the notions of situation, situation inclusion, and of truth at a situation. For example, Grosz and Sidner’s ‘focus space stack’ proposal can be reformulated in terms of situations and their hierarchical organization. For the purposes of this paper, it will be sufficient to know that situations have sub-situations, and that attention can move from a situation to one of its sub-situations, thus restricting the range of objects being considered.

4 The Semantics of Definite Descriptions

I adopt here Heim’s theory of definite descriptions [Heim, 1982].⁴ According to Heim, the interpretation of a definite like *the engine* imposes two constraints on the context. The context has to (i) provide an index for the definite, and (ii) presuppose that the value assigned to that index satisfies the restriction specified by the head noun, i.e., that that object has the property of being an engine.

As far as the first of these constraints is concerned, Heim’s theory can be reformulated in terms of parameters by assuming that part of the meaning of the definite—namely, its referent—is context-dependent, i.e., denotes a parameter d , as follows:

$$(7) \quad \textit{the engine} \rightsquigarrow \lambda P \text{ENGINE}(d) \wedge P(d)$$

Because of the Condition on Discourse Interpretation discussed in §2, having a parameter d in (7) is equivalent to Heim’s requirement that the index assigned to the definite has to be that of an existing file card. I propose to tie the second context-dependent aspect of a definite’s meaning—the condition that context must presuppose that the referent of the definite is an engine—to the ‘search for a discourse topic’ that I discussed in the previous section. I propose, that is, that a listener verifies that the context entails the existence of an object of the type specified by the head noun by identifying a discourse topic that supports the existence of such an object. This can be represented by including a parameter ranging over situations s in the meaning of a definite, such that the situation is (i) part of the common ground and (ii) supports the existence of an object of the right type, as shown in (8):⁵

(8) *the engine* $\rightsquigarrow \lambda P [s \models [\text{ENGINE}(d) \wedge P(d)]]$

5 Visible Situation Definites

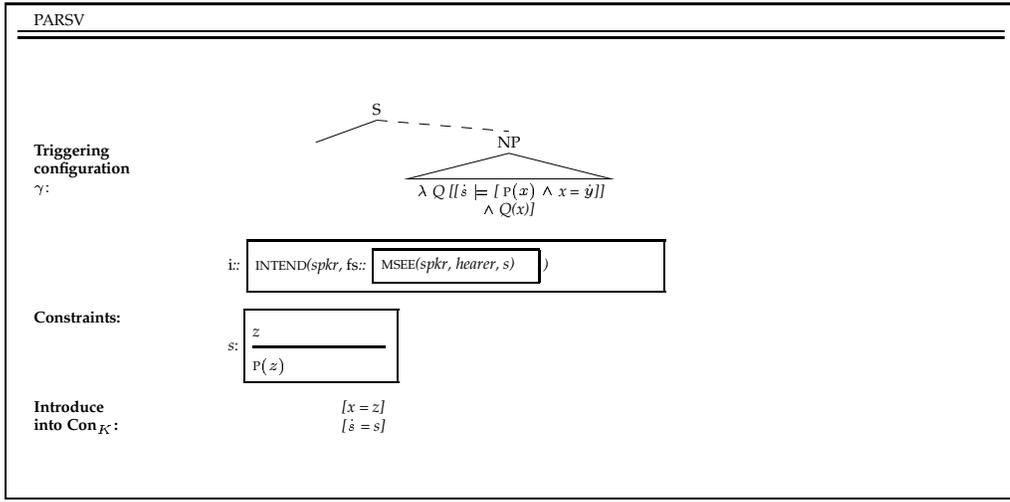
As an example of the way discourse interpretation processes get formalized in the theory presented here, I'll briefly discuss the interpretation of definites interpreted with respect to the *visible situation*. This is the case in which "...the object referred to is visible to both speaker and hearer in the situation of utterance, and is furthermore unique." ([Hawkins, 1978], p.110). An example from the TRAINS conversations is shown below; both conversational participants are looking at a map.

(9) U: while this is happening,
take engine E1 to Dansville,
pick up the boxcar,
and come back to Avon

In the model presented here, visible situation definites are definites that are interpreted with respect to a particular discourse topic, the 'real world,' or one of its sub-situations. It is well accepted [Grosz, 1977] that in order to model the visible situation use, we need to represent the fact that the speaker's attention is focused at certain times on some objects. I assume that this *focus of attention* is a sub-situation of a larger situation that we may call 'real world'. The discourse interpretation principle that leads to the generation of the hypothesis that the definite description *the boxcar* in (9) may refer to a boxcar in the visible situation can be paraphrased as follows:

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

This principle can be formulated as a construction rule in the format proposed by Kamp and Reyle as follows (see [Poesio, 1994a], section 6.5.2, for discussion):



I will just remark that the rule will result in one hypothesis being generated for each potential referent in s ; i.e., in a situation in which s contains two objects of type P , two conflicting hypotheses will be generated, and an ambiguity perceived, unless commonsense knowledge can order the hypotheses.

We are left with the task of specifying how the speaker's intention concerning the mutual focus of attention gets added to the common ground. As observed by Grosz [1977], an attention shift can only be exploited when the participants in a conversation mutually know that the shift took place, on the grounds of some general fact about the conversation. The rules seem to depend on the task at hand; in the TRAINS conversations, the following principle seems to apply:

Follow The Movement : Part of the intended effect of an utterance instructing an agent to move an object from one location to another is to make the terminal location of the movement the new mutual situation of attention.

This principle can similarly be formalized as a construction rule.

I will not discuss in this paper the formalization of the processes involved in the interpretation of anaphoric definite descriptions; this formalization can be found in [Poesio, 1994a]. I will simply remark that in a situation in which a definite description can be interpreted both anaphorically and with respect to the visible situation, the current theory predicts that an ambiguity will be perceived unless one interpretation is more plausible than the other, which seems correct.

Acknowledgments

This work was in part supported by the European Community LRE project 62-051 FRACAS. Thanks to James Allen, David Beaver, Robin Cooper, Janet Hitzeman, David Milward, Len Schubert, Julie Sedivy and David Traum for helpful comments and suggestions.

Notes

¹This corpus [Gross *et al.*, 1993], collected as part of the TRAINS project at the University of Rochester, has served as the basis for most of the work discussed here.

²The abductive model proposed by Hobbs *et al.* [Hobbs *et al.*, 1990] is also formulated in terms of ‘questions to be answered’ by discourse interpretation.

³More precisely, either situations or situation types.

⁴In [Poesio, 1993], a Russellian treatment of definite descriptions was assumed. Some arguments for the adoption of Heim’s theory are given in [Poesio, 1994b].

⁵Reformulated in this way, Heim’s theory becomes very similar to Hawkins’ ‘location theory’ of definite descriptions, according to which the task of the listener upon hearing a definite description is to (i) identify the situation of which the referent is a part, and (ii) identify the referent of the definite within that situation [Hawkins, 1978].

References

- [Barwise and Perry, 1983] Barwise, J. and Perry, J. 1983. *Situations and Attitudes*. The MIT Press.
- [Cooper, 1983] Cooper, R. 1983. *Quantification and Syntactic Theory*. D. Reidel Publishing Company, Dordrecht, Holland.
- [Crain and Steedman, 1985] Crain, S. and Steedman, M. 1985. On not being led up the garden path: the use of context by the psychological syntax processor. In Dowty, D. R.; Karttunen, L.; and Zwicky, A. M., editors 1985, *Natural Language Parsing: Psychological, Computational and Theoretical perspectives*. Cambridge University Press, New York. 320–358.
- [Garrod, 1993] Garrod, S. 1993. Resolving pronouns and other anaphoric devices: The case for diversity in discourse processing. In Clifton, C.; Frazier, L.; and Rayner, K., editors 1993, *Perspectives in Sentence Processing*. Lawrence Erlbaum.
- [Gibson, 1991] Gibson, E. 1991. *A Computational Theory of human linguistic processing: memory limitations and processing breakdown*. Ph.D. Dissertation, Carnegie Mellon University, Pittsburgh.
- [Gross *et al.*, 1993] Gross, D.; Allen, J.; and Traum, D. 1993. The TRAINS 91 dialogues. TRAINS Technical Note 92-1, Computer Science Dept. University of Rochester.
- [Grosz and Sidner, 1986] Grosz, B. J. and Sidner, C. L. 1986. Attention, intention, and the structure of discourse. *Computational Linguistics* 12(3):175–204.
- [Grosz, 1977] Grosz, B. J. 1977. *The Representation and Use of Focus in Dialogue Understanding*. Ph.D. Dissertation, Stanford University.
- [Hawkins, 1978] Hawkins, J. A. 1978. *Definiteness and Indefiniteness*. Croom Helm, London.

- [Heim, 1982] Heim, I. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. Dissertation, University of Massachusetts at Amherst.
- [Hobbs *et al.*, 1990] Hobbs, J. R.; Stickel, M.; Martin, P.; and Edwards, D. 1990. Interpretation as abduction. Technical Note 499, SRI International, Menlo Park, CA.
- [Hwang and Schubert, 1993] Hwang, C. H. and Schubert, L. K. 1993. Episodic logic: A situational logic for natural language processing. In Aczel, P.; Israel, D.; Katagiri, Y.; and Peters, S., editors 1993, *Situation Theory and its Applications*, v.3. CSLI. 303–338.
- [Kamp and Reyle, 1993] Kamp, H. and Reyle, U. 1993. *From Discourse to Logic*. D. Reidel, Dordrecht.
- [Kurtzman and MacDonald, 1993] Kurtzman, H. S. and MacDonald, M. C. 1993. Resolution of quantifier scope ambiguities. *Cognition* 48:243–279.
- [Kurtzman, 1985] Kurtzman, H. 1985. *Studies in Syntactic Ambiguity Resolution*. Ph.D. Dissertation, MIT, Cambridge, MA.
- [Lewis, 1979] Lewis, D. K. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8:339–359.
- [Poesio, 1993] Poesio, M. 1993. A situation-theoretic formalization of definite description interpretation in plan elaboration dialogues. In Aczel, P.; Israel, D.; Katagiri, Y.; and Peters, S., editors 1993, *Situations Theory and its Applications*, vol.3. CSLI, Stanford. chapter 12, 339–374.
- [Poesio, 1994a] Poesio, M. 1994a. *Discourse Interpretation and the Scope of Operators*. Ph.D. Dissertation, University of Rochester, Department of Computer Science, Rochester, NY.
- [Poesio, 1994b] Poesio, M. 1994b. Weak definites. In *Proceedings of the Fourth Conference on Semantics and Linguistic Theory, SALT-4*. Cornell University Press.
- [Raskin, 1985] Raskin, V. 1985. *Semantic Mechanisms of Humor*. D. Reidel, Dordrecht and Boston.
- [Reichman, 1985] Reichman, R. 1985. *Getting Computers to Talk Like You and Me*. The MIT Press, Cambridge, MA.
- [Reiter, 1980] Reiter, R. 1980. A logic for default reasoning. *Artificial Intelligence* 13(1–2):81–132.
- [Seidenberg *et al.*, 1982] 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.
- [Shoham, 1988] Shoham, Y. 1988. *Reasoning About Change*. The MIT Press, Cambridge, MA.
- [Spivey-Knowlton *et al.*, 1994] Spivey-Knowlton, M.; Sedivy, J.; Eberhard, K.; and Tanenhaus, M. 1994. Psycholinguistic study of the interaction between language and vision. In *Proceedings of 12th National Conference on Artificial Intelligence (AAAI-94)*, Seattle.
- [Swinney, 1979] Swinney, D. A. 1979. Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior* 18:545–567.