Context in Abductive Interpretation

Matthew Stone
Computer Science and Cognitive Science
Rutgers University
mdstone@cs.rutgers.edu

Richmond H. Thomason
AI Laboratory
University of Michigan
rich@thomason.org

Abstract

This paper develops a general approach to contextual reasoning in natural language processing. Drawing on the view of natural language interpretation as abduction (Hobbs et al., 1993), we propose that interpretation provides an *explanation* of how an utterance creates a *new discourse context* in which its interpreted content is both *true* and *prominent*. Our framework uses dynamic theories of semantics and pragmatics, formal theories of context, and models of attentional state. We describe and illustrate a Prolog implementation.

1 Problem Statement

Context in discourse and dialogue involves at least two components, *shared information* and *coordinated attention*; each of these components suggests a mechanism through which speakers may frame natural language utterances to fit the context.

- Shared information makes available to the participants a body of facts that characterize the world under discussion (here we mean to be neutral between various characterizations of this availability in terms of common ground, common knowledge, mutual belief, etc.); accordingly, language users may presume that information in an utterance links up with this body of facts, whenever possible (Lascarides et al., 1992; Hobbs et al., 1993).
- Coordinated attention puts certain entities at the forefront of the discussion; accordingly, language users may presume that descriptions in an utterance refer to these entities, whenever possible (Grosz and Sidner, 1986; Grosz et al., 1995; Walker et al., 1997).

How these two potential mechanisms interact is an empirical question—but this question can only be addressed within a framework that describes in precise terms how utterance interpretation interacts with an evolving context of coordinated attention and shared information. Our aim in this paper is to lay out such a framework.

In this brief, initial formulation of the overall theory, we introduce no formal devices that distinguish face-to-face conversation from other genres. Nevertheless, the need for an integrated account of pragmatic context such as ours is clearest in connection with spoken dialogue. Interlocutors in face-to-face dialogue liberally exploit the shared information of their perceptual environment and the coordinated attention of an ongoing real-world collaboration. Empirical studies clearly show speakers drawing on both factors to further efficient communication; for a recent study, see (Brown-Schmidt et al., 2002).

1.1 Interpretation as abduction

On the theory of interpretation as abduction articulated in Hobbs et al. (1993), an interpretation is represented as a proof of the logical form (LF) of the utterance from a background knowledge base (KB) and some further new assumptions. For instance, a telegraphic utterance *lube-oil alarm sounded* might be assigned this LF:

(1.1) $\textit{lube-oil}(o) \;\wedge\; \textit{alarm}(a) \;\wedge\; \textit{nn}(o,a) \;\wedge\; \textit{sound}'(e,a)$

The interpretation might then be represented as the graph in Figure 1, where arrows indicate inferences from KB facts (about a particular quantity of lube oil o_3 and a particular alarm r_5) to LF conjuncts, and a box indicates a literal that is assumed rather than derived. (Think of the variables in the logical forms in Figure 1 as existentially quantified.) A proof in this framework gives a specific way of linking up the meaning of an utterance with the shared information in the context; in Figure 1, for example, the proof spells out how $\textit{for}(r_5, o_3)$, which says that the function of alarm r_5 is to monitor oil o_3, justifies the use of the noun-noun modification structure in the utterance $nn(o, a)$ describing the alarm.

The selection of an interpretation depends on informational preferences, including preferences for merging redundancies, preferences for making certain kinds of assumptions, and preferences for using certain kinds of knowledge. Proofs provide general concrete computational data structures to record such links. They serve as a natural, well structured domain over which to exhibit how these preferences depend on background knowledge. Moreover, computational logic provides increasingly attractive inference algorithms to search for the most preferred interpretation within an interpretation-as-abduction theory (Konrad, 2000; Kohlhase, 2000; Baumgartner and Kühn, 2000; Gardent and Konrad, 2000).

However, Hobbs-style abductive interpretations like the one presented in Figure 1 provide no representation of the changing status of entities in attentional prominence as the context evolves. Take a simple example:

(1.2) She asked her a question.

The LF is as follows.

(1.3) $\mathit{woman}(x) \wedge \mathit{woman}(y) \wedge x \neq y \wedge \mathit{question}(q) \wedge \mathit{ask}'(e', x, y, q)$

The theory offers only two ways that we might single out one woman w_i as the preferred value of x in an abductive proof of (1.3), and so resolve the reference of *she*. We might appeal to knowledge that helps explain why w_i would ask something. Alternatively, following (Lascarides et al., 1992; Hobbs et al., 1993), we might posit further conjuncts to the LF which would require us to infer a rhetorical link between e and the eventualities introduced by prior discourse. But these two strategies are empirically and conceptually problematic: domain information has to be limited and related to the discourse, and LFs need to be specified systematically and compositionally. The obvious rival explanation is simply that w_i is the most salient woman in the context of the utterance, and that *she* refers to the most salient woman (other things being equal). In disallowing this explanation, the interpretation-as-abduction theory cuts off an extremely natural and successful approach to language interpretation.

	KB	LF
	$\mathit{lube\text{-}oil}(o_3) \rightarrow$	$\mathit{lube\text{-}oil}(o)$
	$\mathit{alarm}(r_5) \rightarrow$	$\mathit{alarm}(a)$
$\mathit{for}(r_5, o_3), \mathit{for}(X, Y) \supset \mathit{nn}(Y, X) \rightarrow$		$\mathit{nn}(o, a)$
		$\boxed{\mathit{sound}'(e, a)}$

Figure 1: Interpretation of *lube-oil alarm sounded* in the theory of Hobbs et al. (1993). The proof reflects the substitutions $o = Y = o_3$ (the lube oil) and $a = X = r_5$ (the alarm).

1.2 Context in abductive interpretation

Our proposal is that the interpretation of an utterance is an explanation of how the utterance creates a *new context* in which its content is *both true and prominent*. Thus, we retain an abductive and inferential characterization of interpretation, but we explicitly relativize explanatory goals to particular contexts, and we provide for explicit reasoning about the attentional aspects of those contexts.

Example (1.4) provides a schematic illustration of our approach. We will use the notation $c : g$ to indicate that goal g must be proved in the context associated with term c. To account for the informational and attentional dimensions of the discourse, we use three terms to identify the contexts before and after the utterance: i, a_1 and a_2. Term i describes the informational features of the discourse context as an accumulating body of domain content; the basic LF constrains this context i. Term a_1 identifies the current attentional state of the discourse; for a_1 we prove that y must be *in-focus* (like any pronominal referent), and that x should be particularly central to the discourse (as the referent of a subject pronoun), a property we represent as *in-focus**. (The star records a particular preference for a good explanation of this relationship.) Finally, term a_2 records the new attentional state *after* the utterance is understood; we use the formula $a_1[x < y < q]a_2$ to indicate that a_2 differs from a_1 in that the ranking $x < y < q$ determines the three most salient entities in a_2, with other salient entities from a_1 represented afterwards. Thus an interpretation is an explanation that shows the following:

(1.4) $a_1[x < y < q]a_2 \wedge$
$a_1 : \mathit{in\text{-}focus}^*(x) \wedge a_1 : \mathit{in\text{-}focus}(y) \wedge$
$i : \mathit{woman}(x) \wedge i : \mathit{woman}(y) \wedge x \neq y \wedge$
$i : \mathit{question}(q) \wedge i : \mathit{ask}'(e', x, y, q)$

A proof of (1.4) will spell out how the utterance *she asked her a question* draws on the attentional state a_1 and the shared background i in order to expand the shared background i and create a new attentional state a_2. In the result, i provides the declarative information that she asked her a question, and a_2 places this relationship at the center of attention.

$$
\begin{array}{rcc}
\text{KB} & & \text{LF} \\
 & & \boxed{a_1[x < y < q]a_2} \\
i : \mathit{woman}(s) & \rightarrow & i : \mathit{woman}(x) \\
i : \mathit{woman}(m) & \rightarrow & i : \mathit{woman}(y) \\
 & & \boxed{a_1 : \textit{in-focus}^*(x)} \\
 & & \boxed{a_1 : \textit{in-focus}(y)} \\
s \neq m & \rightarrow & x \neq y \\
 & & \boxed{i : \mathit{question}(q)} \\
 & & \boxed{i : \mathit{ask}'(e', x, y, q)}
\end{array}
$$

Figure 2: Our proposed interpretation for *she asked her a question*. The proof reflects the substitutions $x = s$ (Susan) and $y = m$ (Mary).

Figure 2 shows an explanation that proves this goal for the case where *she* is Susan (s) and *her* is Mary (m); this would count as an interpretation of the utterance on our proposal. We have in such explanations all of the resources of an interpretation-as-abduction theory to describe the role of knowledge in resolving phenomena of local pragmatics. But we can also model the role of attention in interpretation, in terms of preferences for the knowledge to use and the assumptions to make about a_1. And we have a principled way to describe the dynamics of attention across utterances, using the assumptions we make to characterize a_2. Thus we can represent and investigate hypotheses about the interaction of attentional state and shared information in stitching together interpretations.

1.3 Overview

(Hobbs et al., 1993) makes the case for the generality of the abductive approach to discourse using a large and varied inventory of examples: e.g., metonymy, anaphora and ambiguity resolution, compound noun interpretation, and the recovery of discourse relations. But each of these examples has to be solved separately; the framework of that paper does not work well with dialogues or even extended monologues, since there is no way to deal with changes of priorities. In removing this limitation, the dynamic extension of Hobbs' approach presented here provides a framework with the same versatility, but which is capable of giving an integrated analysis for multiple sentences in succession. Work subsequent to (Hobbs et al., 1993) showed that abduction could also be used as a framework for discourse generation; see (Thomason et al., 1996). Although the ideas are illustrated in this paper only with one fairly simple example, we intend them as a general framework for representing and reasoning about not only the interpretation, but the generation of utterances across an evolving context.

In the rest of this paper, we develop this proposal in detail. We first present a theory of discourse context that covers both semantics and pragmatics by specifying the truth of propositions and attentional preferences for resolving ambiguities. The theory draws on theoretical proposals on the role of context in natural language interpretation, such as Stalnaker (1972) and Kaplan (1978); on accounts of the dynamics of discourse context, such as Lewis (1979), Kamp (1981), and Beaver (1992); and on the ideas that have emerged from developments in formalizing context in AI, such as Buvač (1993) and McCarthy and Buvač (1995). Next, we show how to integrate this model of discourse context into the architecture for weighted abduction developed in Stickel (1991) and used in Hobbs et al. (1993), by relativizing axioms, assumptions and goals to contexts, and by determining the plausibility of an explanation as a function of the pragmatic prominence assigned to possible axioms and assumptions in context. As usual for resolution-based proof search systems, the natural realization of our proof rules gives a simple but effective Prolog implementation of this abductive reasoner. Finally, we illustrate the proposal by describing how this framework describes the resolution of pronouns in a complex discourse in which the resolution depends on linguistic constraints, common-sense reasoning, and the evolving prominence of discourse referents.

2 The Representation of Context

Our representation of context must describe both information and attention.[1] Following Stalnaker (1972), we treat shared information in terms of modal operators with a possible-worlds interpretation. A proposition counts as part of the context when it is true in all possible worlds compatible with the discourse participants' mutual knowledge. To characterize the influence of attention, we will follow Kaplan (1978) in creating an abstract disambiguation space; you can think of an *index*, or point in this space, as a simultaneous choice of appropriate values resolving all the ways in which an expression of the language can be ambiguous. If free

[1]The following paragraphs are an informal statement of ideas presented elsewhere. See Thomason (1999) for formal details.

variables are the only source of ambiguity, then an index is simply an assignment of values to variables. If the language contains indexical expressions, like 'I' and 'now', an index will include an evaluation of these expressions. If the language has ambiguous expressions, the index will determine a disambiguation of these expressions. (The values that an index assigns to parameters are intensions; they are functions from possible worlds to various domains.)

In the AI literature on context, a context does two things: it accesses certain axioms or knowledge sources, and it disambiguates ambiguous expressions. If we treat a context c as a pair $\langle [c], i_c \rangle$ consisting of a modal operator $[c]$ and an index i_c, it is natural to revise the usual notation for contexts as follows: $ist(\phi, [c], i_c)$. In this formula, ϕ denotes a function from indices to sets of possible worlds,[2] $[c]$ is interpreted using a relation R_c over possible worlds, and i_c denotes an index. A world w belongs to the the set of worlds assigned to (2) in a model iff for all w' such that wR_cw', w' belongs to the set of worlds assigned to ϕ on disambiguation i_c.

This representation of context is appropriate for tasks in which information that is distributed across a number of knowledge sources must be integrated for reasoning purposes. The information integrator is equipped with its own context, and also knows how other contexts encode information. It receives messages that are labeled according to their sources, and needs to translate them in order to bring them into a single context for subsequent inference.

In natural language interpretation, the disambiguation problem is quite different; there are competing disambiguations of an utterance, the context is unknown, and the problem is to find a plausible disambiguation. In other words, a large part of natural language interpretation is *to infer the intended context*. Naturally, this reasoning uses information concerning what has been said previously and the mutually perceived environment (including features of the utterance like intonation), and naturally this information is often called the "context" of the interpretation. This terminological clash has created much confusion in attempts to use theories of context in connection with natural language. To avoid this confusion, we will use 'S-context' ('Semantic Context') when we have in mind a pair $\langle [c], i_c \rangle$, and will use 'D-context' ('Discourse Context') for the informational situation of the natural language interpreter.

[2]In Kaplan's terms, ϕ denotes a *character*.

A large part of this situation is a changing set of preferences that are used to establish anaphoric references, to resolve ambiguities, and to fill in implicit meanings. We propose, then, to identify a D-context with a triple $\langle [c], I_c, \prec_c \rangle$, where I_c is a set of indices and $\prec_c$ is a partial order over I_c.

The formalization of a linguistic example for abductive interpretation assigns costs to proofs which are associated with various interpretations of an utterance. This can be represented by a D-context that ranks interpretations according to their abductive costs. The abductive search for a least-cost proof is then equivalent to a search for an interpretation that is maximally preferred in the D-context.

Each utterance is not only evaluated with respect to a D-context; it has the potential to change that context. In what follows we will assume that an utterance can change a D-context in two ways: by adding new information to the background for the next utterance and by altering the preferences that apply to disambiguate it.

3 Abduction for D-Context

Our computational realization of this formalism for reasoning in D-context consists of an algorithm for abductive inference in modal logic and a representation for linking abductive modal proofs with preferences for particular interpretations.

3.1 Abductive modal inference

We couch the declarative part of our contextual reasoning in a modal logic; modal operators of the form $[c]$ are associated with contexts; we also have a distinguished operator $\Box$ which we use to specify axioms that are true in all contexts. We assume atomic formulas as in Prolog, using A to schematize them; we also assume an always true atom $\top$. If κ is a sequence of context operators of the form $[c_0] \ldots [c_n]$ (possibly empty), we use the notation $\kappa(\phi)$ to name the formula $[c_0] \ldots [c_n]\phi$; we use $\kappa \circ \kappa'$ to denote the concatenation of κ and κ'.

We use a fragment of modal logic that extends Prolog in a simple way. Goals G are modalized atomic formulas; clauses P are modalized Prolog clauses whose antecedent and conclusion formulas may themselves be modalized:

$$\begin{aligned} G &::= \kappa(A) \mid \top \\ P &::= \kappa'(\kappa_1(A_1) \ldots \kappa_m(A_m) \to \kappa''(H)) \mid \\ &\quad\;\; \Box\kappa'(\kappa_1(A_1) \ldots \kappa_m(A_m) \to \kappa''(H)) \end{aligned}$$

The interpretation of this language is the usual Kripke semantics, with each $[c]$ treated as a K modality and $\Box$ treated as an S4 modality with an accessibility relation containing that of each $[c]$; see e.g., Fitting and Mendelsohn (1998).

In this fragment, we can streamline the prefixed tableau inference method of Fitting and Mendelsohn (1998) with simplified representations of possible worlds suitable for reasoning with definite clauses. Each *query* can be simplified to the form $\kappa : A$, where κ is a modality and A is atomic formula. In resolution, we must match the head H of our clause with the query atom A, but we must also match the context κ of the query with the context $\kappa' \circ \kappa''$ or $\Box \circ \kappa' \circ \kappa$ of our clause ($\Box$ goes proxy for any modality κ^*). Baldoni et al. (1998) show that this strategy suffices for sound and complete inference in this modal fragment.

To refine this procedure into an abductive inference algorithm, we introduce *marked* queries, which consist of a query together with a designation *resolved*, *assumed* or *unsolved*; now, as Stickel does, we can describe a partially-constructed abductive proof by a list of each marked query derived and its current proof status. In this list, a *resolved* or *assumed* query all of whose predecessors are also *resolved* or *assumed* has in fact been proved (abductively) from the common premises. This justifies a new factoring rule to derive a subsequent query by reusing this proof.

The resulting algorithm constructs abductive explanations by applying any of four rules nondeterministically to transform one list of marked queries into another until all the formulas in the list are either resolved or assumed. These rules can be implemented immediately as Prolog clauses, by using Prolog variables and unification to represent object-level variables in formulas and by using Prolog search to manage nondeterminism in the application of rules. (In our implementation, we have gone beyond this only in testing for conflicts, in assigning costs to proofs, and in using bounds on cost to impose an iterative-deepening search regime.)

Suppose the current state is represented as a list $L = Q_1, \ldots, Q_n$ with $Q_i = \kappa : A$ is the first (leftmost) query marked *unsolved*. We can transform the state as follows:

- *Truth.* If A is $\top$, derive a new state exactly like L except that the label of Q_i is *resolved* rather than *unsolved*.
- *Assumption.* Derive a new state exactly like L except that the label of Q_i is *assumed* rather than *unsolved*.
- *Resolution.* Select a clause R of the form P or $\Box P$, where P is
$$\kappa'(\kappa_1(A_1) \ldots \kappa_m(A_m) \to \kappa''(H))$$
with its variables renamed, if necessary, so that it has no variables in common with L. Suppose H and A are unifiable with most general unifier σ, $\kappa = \kappa^* \circ \kappa' \circ \kappa''$, and unless R is $\Box P$ then κ^* is empty. Then derive the new state:
$$Q_1\sigma, \ldots, Q_{i-1}\sigma,$$
$$\kappa^* \circ \kappa' \circ \kappa_1 : A_1\sigma[\mathit{unsolved}], \ldots,$$
$$\kappa^* \circ \kappa' \circ \kappa_m : A_m\sigma[\mathit{unsolved}],$$
$$\kappa : A\sigma[\mathit{resolved}], \ldots, Q_n\sigma$$
- *Factoring.* Suppose some element Q_s describing a query $\kappa : H$ precedes Q_i in L, and H and A are unifiable with most general unifier σ. Then derive a new state suppressing the duplicate proof of Q_i:
$$Q_1\sigma, \ldots, Q_{i-1}\sigma, Q_{i+1}\sigma, \ldots, Q_n\sigma$$

An abductive proof for a query Q is a sequence of lists whose initial element is $Q[\mathit{unsolved}]$, in which each later list is obtained from its predecessor by the application of one of these transformations, and which terminates in a list none of whose elements are marked *unsolved*. An abductive proof determines an *answer* to the query, a pair $\langle Q_0, \Delta\rangle$ consisting of the found instantiation Q_0 of Q (determined from the final element of the terminal list) together with the postulated assumptions Δ (a multiset consisting of a formula $\kappa(A)$ for each element of the form $\kappa : A[\mathit{assumed}]$ in the terminal list).

Space limitations prevent us from describing the workings of the algorithm and correctness results in further detail here.

3.2 Preferences in D-contexts

Any nontrivial example of discourse provides alternative interpretations; we therefore need preferences in order to choose among alternative explanations. Stickel treats these overall preferences as a function of the assumptions and axioms used in a proof. We generalize this idea to provide for the dependence of these preferences on a dynamically evolving discourse context.

Assumption costs are specified indirectly by propagating values through the proof. The initial query is annotated directly with an assumption cost, while costs for further queries are determined by annotating rules so that each subgoal G_j is associated with an *assumability function* f_j:

$$(\Box)\kappa'(G_1 / f_1 \wedge \ldots \wedge G_m / f_m \to \kappa''(Q))$$

When such a rule is instantiated to values x and resolved against a query with cost c in a context $\kappa^* \circ \kappa' \circ \kappa''$, each new subgoal G_j determines a query with an assumption cost calculated as

$$c \times f_j(\kappa^* \circ \kappa', x)$$

Each axiom is also annotated with a function f_r, where $f_r(\kappa^* \circ \kappa', x)$ is a cost for the use of the axiom in the explanation. To get the overall cost for an answer $\langle Q, \Delta \rangle$, we sum the assumption costs associated with the elements in Δ, together with axiom costs for each resolution step in the proof.

To illustrate the context-sensitivity of assumption costs, take the familiar (nonlinguistic) case of the WET-LAWN, which can be explained either by RAIN or by SPRINKLER. The relevant rules are these.

(3.1) $\Box(\text{SPRINKLER} \,/\, f_1 \;\rightarrow \text{WET-LAWN})$
(3.2) $\Box(\text{RAIN} \,/\, f_2 \;\rightarrow \text{WET-LAWN})$

Here, we represent the costs associated with using rules (3.1) and (3.2) as functions $f_1(\kappa)$ and $f_2(\kappa)$ of the reasoning context. Thus, even though we allow these axioms to apply in all contexts, in a context κ where $f_1(\kappa)$ is relatively high and $f_2(\kappa)$ relatively low, (3.2) will lead to RAIN as the preferred explanation of WET-LAWN. When $f_1(\kappa)$ is relatively low and $f_2(\kappa)$ is relatively high, SPRINKLER will be the preferred explanation.

In discourse interpretation, it is vital not only that assumability functions take different values across contexts, but also that those values can be established dynamically, as each sentence draws on the pragmatic state set up by the interpretation of its predecessors. To enable this, we provide the abductive reasoner with a high-level specification of the behavior of assumability functions in new contexts—an abstract characterization of the discourse preference dynamics.

4 An Example

We will use anaphora resolution to illustrate the approach to context in natural language interpretation articulated in the previous sections. Consider the following three-sentence text.

(4.1) Susan met Mary.
She asked her a question.
She answered no.

On the natural interpretation, it is Susan who asked Mary a question, and Mary who answered no.[3]

In our model, this interpretation is an explanation of how discourse (4.1) updates an evolving context. As with (1.4), we describe this in terms of a term i for the information in the discourse and terms a_0, a_1 and a_2 for the changing attentional state of the discourse. From the first sentence, we learn that Susan met Mary, and update attention so that Susan and Mary are salient, by proving

(4.2) $a_0[u < v]a_1 \wedge i : \mathit{susan}(u) \wedge i : \mathit{mary}(v) \wedge i : \mathit{meet}'(e, u, v)$

The second sentence is analyzed as (1.4); the third is analyzed along similar lines:

(4.3) $a_2[z]a_3 \wedge a_2 : \mathit{in\text{-}focus}^*(z) \wedge i : \mathit{woman}(z) \wedge i : \mathit{answer}'(e'', z, q', \mathit{no})$

Figure 3 diagrams a proof corresponding to the intended interpretation of this discourse. Note how the commonsense inference (4.4) helps to explain the answer.

(4.4) If an event E in which X asks Y something has just occurred, an event E' in which Y gives X the answer *no* is a strong possibility; this overrides salience considerations based on grammatical relations only.

By using this in the proof of Figure 3, we need only assume a Hobbs-style default $etc_{q-a}(e', e'')$.

With such representations, it is straightforward to describe general criteria which favor the interpretation of Figure 3 over all alternatives, and to realize those criteria as preferences so that our abductive theorem prover automatically identifies the intended interpretation for (4.1). Our criteria (4.5) and (4.6) correspond to the forward-looking center ranking of Centering Theory and its preference for continue transitions in discourse (Walker et al., 1997).

(4.5) A sentence with a feminine subject and direct object induces a context in which the subject is most ‘she’-salient (i.e., resolutions of ‘she’ to the subject take a lower cost in the induced context) than the direct object, and the direct object is more ‘she’-salient than anything else.

(4.6) An interpretation in which the subject refers to a more salient entity is to be preferred to one where it refers to a less salient one (in the absence of more compelling considerations).

[3]Although (4.1) involves only a single speaker, the same tradeoffs in interpretation appear in dialogue, as this example illustrates:

A: Did Susan meet Mary? B: Yes. A: Did she invite her? B: Yes. And she accepted.

With rules to deal with the context-changing effects of questions, our techniques could formalize this reasoning, too.

$$
\begin{array}{lccc}
\text{KB} & \text{LF 1} & \text{LF 2} & \text{LF 3} \\
 & \boxed{a_0[u < v]a_1}^{1} & \boxed{a_1[x < y < q]a_2}^{2} & \boxed{a_2[z]a_3}^{3} \\
i : \mathit{susan}(s) \quad \rightarrow & i : \mathit{susan}(u) & & \\
i : \mathit{mary}(m) \quad \rightarrow & i : \mathit{mary}(v) & & \\
 & \boxed{i : \mathit{meet}'(e, u, v)} & & \\
 & & \boxed{a_1 : \mathit{in\text{-}focus}^{*}(x)}_{1} & \\
 & & \boxed{a_1 : \mathit{in\text{-}focus}(y)}_{1} & \\
i : \mathit{woman}(s) & \rightarrow & i : \mathit{woman}(x) & \\
i : \mathit{woman}(m) & \rightarrow & i : \mathit{woman}(y) & \\
s \neq m & \rightarrow & x \neq y & \\
 & & \boxed{i : \mathit{question}(q)} & \\
 & & & \boxed{a_2 : \mathit{in\text{-}focus}^{*}(z)}_{2} \\
i : \mathit{woman}(m) & & \rightarrow & i : \mathit{woman}(z) \\
 & & \boxed{i : \mathit{ask}'(e', x, y, q)} \quad \rightarrow & \\
\multicolumn{3}{l}{i : \mathit{ask}'(E, X, Y, Q) \wedge \boxed{\mathit{etc}_{q\text{-}a}(E, E')} \supset \mathit{answer}'(E', Y, Q, \mathit{no}) \quad \rightarrow} & i : \mathit{answer}'(e'', z, q', \mathit{no})
\end{array}
$$

Figure 3: Figure 3: Intended abductive interpretation for discourse (4.1). Subscripts indicate assumptions whose costs depend on the correspondingly superscripted context description. The proof reflects the substitutions $u = x = X = s$ (Susan), $v = y = Y = z = m$ (Mary), $Q = q = q'$ (the question), $E = e'$ (the asking) and $E' = e''$ (the answering).

They yield to the commonsense criterion of (4.4).

Abductive proof search explores a space of interpretations in which the variables x, y and z are unified either to s or to m. The resolution of x and y in the second sentence is independent of that of z in the third; so let's begin with the second. Inconsistency eliminates the possibilities where $x = y$. Now because by (4.5) s is more salient than m in a_1, we get different cost-factors $w_{x=s} = w_{y=s} < w_{x=m} = w_{y=m}$ for different *in-focus* assumptions. Meanwhile, the goals created according to (4.4) assign a cost u_x to assuming *in-focus*$^*(x)$ for the subject. This exceeds the cost u_y for assuming *in-focus*(y) for the object: $u_y < u_x$. The overall cost is therefore minimized when we resolve x to Susan and y to Mary (at $w_{x=s} \times u_x + w_{y=m} \times u_y$). Turning now to the resolution of z, the context contains an event of Susan asking Mary a question. The key alternative here is whether we take z to be Susan, assuming the answering outright with some cost u_c, or we take z to be Mary, using the asking event to derive Mary's answer with reduced cost u_m. According to (4.4) and (4.6), the decrease in cost from u_c to u_m outweighs the additional cost of chosing Mary, the less salient 'she', for z.

5 Conclusions and Future Work

The changing interpretation of pronouns against an evolving semantic and pragmatic context in discourse has been studied from many linguistic and computational perspectives. The relation of the more computationally oriented work, such as Centering Theory (Walker et al., 1997), to general theories of context is not entirely clear; and the literature provides no general mechanism for integrating reasoning based on linguistic structure and pragmatic context with other information to capture coherence or commonsense inferences. On the other hand, treatments of anaphora resolution couched in terms of semantic approaches such as the Structured Discourse Representation Theory used in Lascarides et al. (1992), account only for the connection between semantics, world knowledge and discourse coherence. In place of general concepts of salience, they offer a taxonomy of specific rhetorical and informational connections that justify apparently salient interpretations—which can seem quite cumbersome. Recent extensions of the interpretation-as-abduction theory (Konrad, 2000; Kohlhase, 2000; Baumgartner and Kühn, 2000; Gardent and Konrad, 2000) seem to suffer from the same drawbacks. The problem is compounded in Hobbs et al. (1993), where context change is not properly represented at all. Our treatment of pronoun resolution in (4.1) allows different kinds of information to be specified simply and separately, but to be combined straightforwardly in reasoning about a changing context.

We believe that the general approach we have articulated and formalized in this paper should enable a theoretically informed investigation of tradeoffs between information and attention in discourse. Promising cases for formalization include the information-dependent models of attention of Kehler (2001), where inferred rhetorical connections change but do not eliminate attentional preferences; and context-dependent representations of apparently inconsistent utterances, like Strawson's (1952) *He didn't jump*, in which a speaker's appeal to coordinated attention requires us to understand an utterance as rejecting claims we would otherwise have taken as shared.

6 Acknowledgments

This material is based on work supported by the National Science Foundation under Grants No. IRI–9314961, "Integrated techniques for natural language generation and interpretation" and 9818322, "A laboratory for interactive applications of computational vision and language".

References

Matteo Baldoni, Laura Giordano, and Alberto Martelli. 1998. A modal extension of logic programming: Modularity, beliefs and hypothetical reasoning. *Journal of Logic and Computation*, 8(5):597–635.

Peter Baumgartner and Michael Kühn. 2000. Abducing coreference by model construction. *Journal of Language and Computation*, 2(1):175–190.

David Beaver. 1992. The kinematics of presupposition. Technical Report LP–92–05, ILLC, University of Amsterdam.

Sarah Brown-Schmidt, Ellen Campana, and Michael K. Tanenhaus. 2002. Reference resolution in the wild: On-line circumscription of referential domains in a natural interactive problem-solving task. In *Proceedings of the Cogntive Science Society*.

Saša Buvač. 1993. Propositional logic of context. In *Proceedings of AAAI*, pages 412–419.

Melvin Fitting and Richard L. Mendelsohn. 1998. *First-Order Modal Logic*. Kluwer.

Claire Gardent and Karsten Konrad. 2000. Interpreting definites using model generation. *Journal of Language and Computation*, 1(2):175–190.

Barbara J. Grosz and Candice L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12:175–204.

Barbara J. Grosz, Arivind Joshi, and Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):227–253.

Jerry Hobbs, Mark Stickel, Douglas Appelt, and Paul Martin. 1993. Interpretation as abduction. *Artificial Intelligence*, 63(1–2):69–142.

Hans Kamp. 1981. A theory of truth and semantic representation. In J.A.G. Groenendijk, T.M.V. Janssen, and M.B.J. Stokhof, editors, *Formal Methods in the Study of Language*. Mathematisch Centrum, Amsterdam.

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

Andrew Kehler. 2001. *Coherence, Reference and the Theory of Grammar*. CSLI.

Michael Kohlhase. 2000. Model generation for discourse representation theory. In *14th European Conference on Artificial Intelligence*.

Karsten Konrad. 2000. *Model Generation for Natural Language Interpretation and Analysis*. Ph.D. thesis, Universität des Saarlandes. Available at http://www.ags.uni-sb.de/˜konrad.

Alex Lascarides, Nicholas Asher, and Jon Oberlander. 1992. Inferring discourse relations in context. In *Proceedings of ACL*, pages 1–8.

David K. Lewis. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic*, 8(3):339–359.

John McCarthy and Saša Buvač. 1995. Formalizing context (expanded notes). Available from http://www-formal.stanford.edu/buvac.

Robert C. Stalnaker. 1972. Pragmatics. In Gilbert Harman and Donald Davidson, editors, *Semantics of Natural Language*, pages 380–397. D. Reidel.

Mark Stickel. 1991. A prolog technology theorem prover: a new exposition and implementation in prolog. Technical Report 464, SRI International.

P. F. Strawson. 1952. *Introduction to Logical Theory*. Methuen.

Richmond Thomason, Jerry Hobbs, and Johanna Moore. 1996. Communicative goals. In Kristiina Jokinen, Mark Maybury, and Ingrid Zukerman, editors, *Proceedings of the ECAI 96 Workshop on Gaps and Bridges: New Directions in Planning and Natural Language Generation*. Springer.

Richmond H. Thomason. 1999. Type theoretic foundations for context, part 1: Contexts as complex type-theoretic objects. In Paolo Bouquet et al., editors, *Modeling and Using Contexts: Proceedings of the Second International and Interdisciplinary Conference, CONTEXT'99*, pages 352–374. Springer.

Marilyn A. Walker, Arivind K. Joshi, and Ellen Prince, editors. 1997. *Centering Theory in Discourse*. Oxford University Press.