Mental models and pragmatics: the case of presuppositions

Guido Boella, Rossana Damiano and Leonardo Lesmo
Dipartimento di Informatica e Centro di Scienza Cognitiva
Università di Torino
{guido,rossana,lesmo}@di.unito.it

Keywords: Mental models, linguistics, pragmatics and presuppositions

Abstract

We claim mental models are a framework that allows to shed light on the phenomenon of presuppositions. A plan-based lexical representation for verbs, together with the effect of conversational implicatures that discharge possible mental models, are the key features of this proposal.

Introduction

Presuppositions are what the utterance of a sentence assumes to be true (or takes for granted). They can be triggered by different elements in sentences, going from the use of definite NP's (*the King of France is bald* presupposes *there is a (unique) King of France*) to the presence of factive verbs (*he regrets that he has been impolite* presupposes *he has been impolite*). These inferences are characterized by resistance to negation (e.g. *he does not regret that he has been impolite* presupposes *he has been impolite*), and by cancellability in the presence of contextual information.

This paper addresses what may be called "lexical" presuppositions. In particular, the event structure, as described by certain verbs, allows to draw some inferences about the described situation. These verbs are classified in (Beaver, 1997) as "signifiers of actions and temporal/aspectual modifiers": "most verbs signifying actions carry presuppositions that the preconditions for the action are met" (Beaver, 1997) page 944.

We think that what accounts for "preconditions for the action" still requires a more precise characterization. Where do preconditions come from? It seems rather vague to state that *leaving* is a precondition for *arriving* and *climbing* is a precondition for *reaching the top*. We claim that these presuppositions arise naturally from a representation of the semantics of verbs in terms of actions and plans describing the steps required to carry out the activity referred to by the verb. It may be observed that a plan-based representation is complex, and rather expensive to build. However, we have noticed elsewhere that it is independently required both for accounting for the meaning of communication verbs (Goy and Lesmo, 1997) and for the maintenance of coherence in dialogues (Ardissono et al., 1998).

The basic idea is that, in order to understand an utterance, one must build a mental representation of the described situation: it is clear, e.g., that any representation of *arrive* must include that of *leave*. The role played by *mental models* (Johnson-Laird, 1983) is to provide a reasoning framework for these representations, that allows to explain why presuppositions survive in certain contexts. For instance, these representations make it possible to account for the fact that *John didn't arrive* presupposes *John left*, as predicted by the projection of presuppositions across negation; moreover, this inference is correctly modeled as a defeasible one, since it is possible to say *John didn't arrive, he didn't even leave*.

The interpretation is carried out in the following way: first of all, the aspectual features are accounted for by means of a plan-based representation of the lexical meaning; moreover, the temporal features represented by mental models in the style of (Schaeken et al., 1996), as well as the interpretation of the negation and of the context; this amount to building one of more *mental models* representing the described event. The information shared by all models is what the sentence entails.
At this point, potential conversational implicatures are applied: their effect is to discharge those mental models of the event that could have been more precisely described by alternative linguistic expressions. In this way, we gain more information, since we are left with the smaller set of mental models: presuppositions are the information that becomes shared by all these remaining models.
However, implicature can be defeated in the presence of further contextual information, thus blocking the discharge of models: presuppositions cannot arise anymore, because they do not occur in all event models, therefore appearing defeasible too.

In the following, we will describe a plan-based semantic representation of action verbs and our treatment of negation and of the temporal expression *before*; then, the conversational implicature mechanism will be introduced. Next, we face the problem of the projection of presuppositions. Comparison with related approaches and conclusions close the paper.

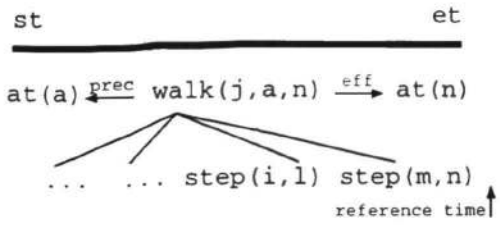

Figure 1: The mental model of *John arrived to* n. The st-et line represents the event time of the action.

A plan-based representation of verbs

We will mainly consider verbs denoting actions that are intentionally executed by an agent. The meaning of these verbs is represented by action schemata, that describe how actions are carried out by a sequence of steps. When a sentence is interpreted, an action instance is built on the basis of the schema corresponding to the main verb; then, a set of constraints expressing the temporal and aspectual information conveyed by the other linguistic elements (aspectual predicates, adverbials, verb arguments and so on) is added to this representation, by means of condition-action rules (Boella and Damiano, 1999). Temporal constraints refer to the occurrence of the action instance with respect to speech time and reference time, following a reichenbachian temporal reference schema.

An action schema Act is composed of arguments, among which the agent (agt (Act)) and the start and end time of the action (st(Act) and et(Act), respectively, denoting temporal points), the preconditions and effects of the action, and the action decomposition (body(Act)), composed of steps; the start and end time of the steps can be specified too.

Usually, a mental model representing an action does not contain all the step instances (tokens) of the plan but only a subset of them. Since steps express the focused part of the action and its temporal placement, only the steps must be included that are needed to represent the constraints resulting from the interpretation. The remaining steps can be later inferred and added to the representation, if they become necessary for reasoning purposes. Moreover, action schemata can be only partially instantiated, to account for the fact that linguistic expressions describe an event by highlighting only certain features of its, that the speaker considers relevant to his goals. In our approach, this corresponds to building models in which only the currently relevant steps are represented, in order to focus on a specific phase of the action or to represent the fact that the action has been only partially executed.

For example, the interpretation of *John walked to the store* contains only the first and the last steps of the plan and constrains them to precede the reference time (here coinciding with the speech time); on the other hand, *John arrived* is not represented as a punctual event but as an instance of the action *move* (that, in this context, specializes into *walk*) containing in its decomposition only the last steps of the plan, plus the temporal constraints specifying that both the whole action of walking and its last steps happened before the reference time (in Figure 1 the decomposition links are denoted by the lines connecting the walk(j,a,n) token to the tokens representing the steps). On the contrary, if the sentence to be interpreted were *John was arriving*, an instance would be built where the reference time occurs within the sequence of steps that conclude the action. In both cases, the steps preceding the final sequence are not represented, but are assumed to exist on the basis of the action instance inner relations and can be consequently added to the representation if they become relevant. According to this representation, the mental model of *John arrived* entails the model of *John left*, where, in turn, the initial steps of the action *walk* are constrained to precede the reference time.

Therefore, the presupposition *John left* is not confined to a separate set of propositions that must be accommodated in the representation as in (Van Der Sandt, 1992); moreover, the requirement is satisfied that presuppositions are in some sense inserted at the beginning of the sentence interpretation (Beaver, 1997) and not added or calculated afterwards, as in (Karttunen, 1974; Gazdar, 1979b).

On the other hand, we do not face here the problem of the anaphorical character of presuppositions, exemplified in *he left an hour ago but he didn't arrive*, where the two clauses refer to two phases of the same underlying action of walking.

The representation of negation

We have adopted a peculiar treatment of negation, in order to account for the fact that the negated event was somehow expected to happen. We interpret such expectations by representing them in terms of intentions attributed to the involved agent; mental models containing unexecuted action instances can be readily interpreted as an agent's mental description of another agent's intentional state (Bratman et al., 1988).

The representation of *John did not arrive* includes the related plan instance of walking, and its decomposition into steps; the negation is not represented by labeling the last steps in the plan instance as denied (as mental model theory prescribes (Johnson-Laird, 1983)). On the contrary, starting from the premise that the walking action did not end before the reference time R[1], all the representations that include the plan instance (act) and satisfy the constraint et(act) $\geq$ R are allowed. Note that

[1] Otherwise we would have the representation of *John arrived*.

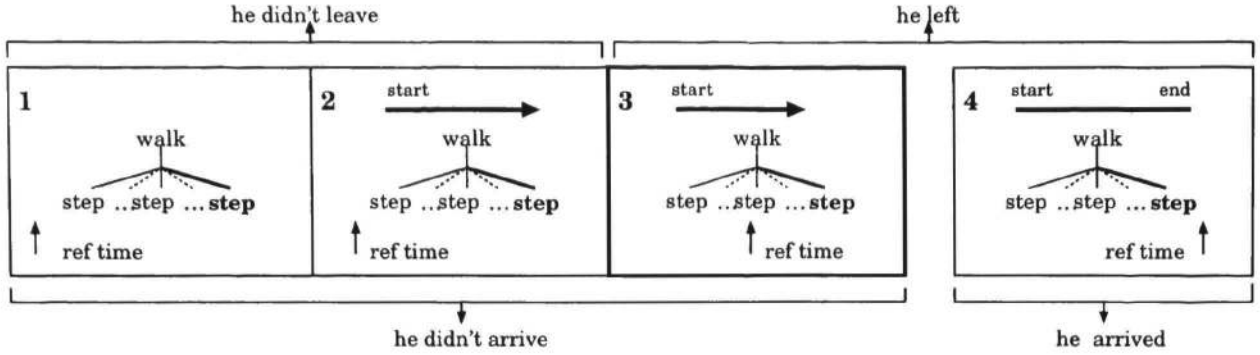

Figure 2: The relations among the mental models concerning *walk* (the thick line show whether the steps have been actually performed).

this does not imply that the *walk* action will necessarily be completed, i.e. that John will arrive later.

By using mental models, it immediately becomes apparent that two pieces of information have to be included in any event description: the fact that the end of the action necessarily follows the reference time and, given the internal constraints of the action schema, the fact that the beginning of the action precedes its conclusion. Given these constraints, two models are possible (Schaeken et al., 1996): in the former the whole action follows the reference time (R), in the latter some steps of the action precede it:

premises			models		
R	et(act)	}	st(act)	R	et(act)
st(act)	et(act)	}	R	st(act)	et(act)

Actually, we have to add another variable to the model: whether the action is conceived as (partially) executed or not. This question concerns only the second model,since the part of an action preceding the reference time has certainly been executed: either the action has been or will be executed after the reference time or it hasn't and it remains as a representation of the intention of the agent.

So, the interpretation of *John did not arrive* consists of the first three models in Figure 2.

Our representation of negative sentences is strictly related to the meaning of the conjunction *before*. *Before* does not imply that the action in the subordinate clause actually happened, as shown by *Mozart died before finishing his Requiem*. As in case of negative sentences, it simply states an expectation (again represented as intentions): that the *finish* event was expected to occur at a time which follows the death (d):

d	st(act)	et(act)
st(act)	d	et(act)

Note that, for this reason, *before* is not symmetric with respect to *after*, contrarily from what (Asher and Lascarides, 1998) claim, since the latter entails the execution of both related actions: *Salieri finished Mozart's Requiem after he died in 1791*.

The parallel between negation and *before* for what concerns presuppositions will be examined below.

Conversational implicature

The gricean notion of conversational implicature is exploited to explain why some of the models are later discharged. In particular, we will consider a particular inference licensed by the maxim of quantity, the scalar implicature (Gazdar, 1979a).

When an item belongs to a salience order (i.e. a scale), the fact that the speaker has used this item instead of a different one in the scale causes the hearer to draw the inference that the speaker was not in the position to use any of the higher elements of the scale (see the model of (Hirschberg, 1985)).

For example, from *some composers died young* it is possible to infer *not all composers died young*, even if *some* per se does not entail *not all*. But *some* belongs to the scale *one, a few, some, many, most of, all*, where higher elements entail the lower ones. If the speaker has used *some* and, at the same time, he is assumed to respect the gricean maxim of quantity, this means that he cannot use the stronger element *all*, and, therefore, he intended to say that some composers last longer.

In our framework, this means that an expression (e.g. *some*) is initially interpreted by means of a set of mental models and, if some of them correspond exactly to the interpretation of another expression (e.g. *all*), then they are discharged:

Some A are B	All A are B	
A		
A – B		
A – B		
B		
A – B	A – B	}
A – B	A – B	} Models discharged by
B	B	} the scalar implica-
A – B	A – B	} ture
A – B	A – B	}

Arrive belongs to a scale with respect to *leave* and the same holds for *finish* and *begin*, since in both cases the

former items entail the latter ones; in fact, the beginning of an action is present in every model representing its conclusion.

By reasoning with mental models it is apparent why, in case of negative sentences, scales (e.g. *leave, arrive*) can be reversed (*not arrive, not leave*). The interpretation of *John did not arrive* produces three mental models (see Figure 2); two of them (1 and 2) also represent the interpretation of *John did not leave*, as only in (3) st(act) $\leq$ R.

But, to describe them, the speaker would have more appropriately used *not leave*, a higher item in the scale, therefore, we can discharge the first two models from the interpretation, and keep the last one.
The presupposition *John left* now emerges since it is contained in all the remaining mental models (3).

Note that we are assuming that the speaker knows whether the higher elements of the scale (*not leave*) hold or not, otherwise this inference would not be possible. Anyway, the hearer usually knows whether the speaker has this kind of knowledge, thanks to the context in which the sentence is uttered.

In this way, the conversational implicature mechanism produces the presupposition, though in a rather indirect way, by deleting the mental models that can be better described by other sentences. Conversational implicatures are a defeasible kind of inferences: in presence of contextual information they may disappear. In our case, the cancellation of the implicature prevents the removal of the two mental models representing the fact that John has not left; therefore the presupposition cannot be drawn anymore. In *John did not arrive since he did not leave* the second clause re-asserts the first two models of Figure 2, while it negates the third one (3) that implies that John left. In this way, the basis for the conversational implicature does not hold anymore since both items *arrive* and *leave* are denied.
In case of positive sentences, like *John arrived*, the presupposition becomes an entailment, since it is contained from the beginning of the interpretation process in the only mental model representing the sentence interpretation (4 in Figure 2), and cannot be cancelled (**John arrived but he didn't leave*).

Sentences involving *before* undergo the same reasoning based on conversational implicatures. From *Mozart met Casanova before finishing his Don Giovanni*, one can infer that he met Casanova in the period in which he was writing his opera. As stated above, *before* allows the construction of two models. One model represents the interpretation of *before he started writing his Don Giovanni*, but, since the speaker didn't use this sentence, this model is discharged, and the only model left is the one where the writing action contains the meeting event.

Furthermore, from this model it is possible to draw a further inference: Mozart actually finished his work. In the mental model we have no explicit information about the conclusion of the action of writing the composition. The inference is then licensed by another kind of motivation; as we stated above, plan instances constitute a description of an agents' intentions and the distinguishing feature of intentions is their persistency (Bratman et al., 1988): if nothing prevents him, the agent will carry out his current intentions. However, this is a defeasible kind of reasoning, and, if more information is added, the inference will be canceled: see *Mozart died before finishing his Requiem.*

The projection problem

The projection problem consists in explaining when and why the presuppositions of a clause become the presuppositions of the whole sentence where it occurs.
We start from the projection problem in disjunctive sentences. *John has stopped smoking* not only presupposes *John smoked* but, actually, implies it. In fact, the interpretation of the aspectual predicate *stop* consists of a process of smoking to which it is added the fact that the agent has not been smoking for a given period of time (see (Boella and Damiano, 1999)).
Nevertheless, in the sentence *either John stopped smoking or he never smoked* this presupposition does not hold anymore. However, since *John smoked* is implied by the first clause, we cannot resort to the cancellability of presuppositions.
A disjunction of clauses A $\vee$ B is represented by three mental models:[2]

A B
$\neg$A B
A $\neg$B

By substituting A with the interpretation of *John stopped smoking* and B for that of *John never smoked*, we obtain a set of potential situations. The interpretation of the negated clause $\neg$A consists of some mental models, to which the interpretations of the clause *B* are added, resulting in a set of integrated models; then, the conversational implicatures are applied; finally, the inconsistent models are discharged (e.g. those in which is asserted that John smoked and never smoked).

John stopped smoking and *never smoked*
John did not stop smoking and *never smoked*
John stopped smoking and *smoked*

Now we add the entailments of the asserted first clause and the presuppositions of the negated one (underlined):

John stopped smoking and smoked and *never smoked*
John did not stop smoking and <u>smoked</u> and *never smoked*

[2] We directly flesh out the explicit models of the disjunction: in principle one should start from the implicit model that contains only the positive information:

A
B

But in our example the negation conveys positive information, i.e. the expectation that the action happens.

John stopped smoking and smoked and *smoked*

The first model is clearly contradictory and is discharged. On the contrary, in the model ¬A B the presupposition arising from a the negated disjunct has a defeasible character, so it is canceled and the model kept. The third model is fine. At the end of the interpretation process, we have the following mental models:

John did not stop smoking and *never smoked*
John stopped smoking and *smoked*

Does this representation imply *John smoked*? Certainly not, as the two models contain opposite information.
On the contrary, in a sentence like *either John stopped smoking or he is now ill* the presupposition that John smoked is projected from the first clause to the whole sentence. In fact, the interpretation produces the following consistent models:

John stopped smoking and smoked and *is ill*
John did not stop smoking and smoked and *is ill*
John stopped smoking and smoked and *is not ill*

Now, all the models contain the information that John smoked, so, from the disjunctive sentence, it is possible to draw the inference that John smoked.
Note that the presuppositions of the two clauses are independently added to the interpretation of the whole sentence and the single models containing them are discharged if inconsistent. Therefore, differently from (Karttunen, 1974)'s approach, we are able to cope with cases in which the two clauses convey contradictory presuppositions as in: *either Fred knows he's won or he's upset that he hasn't* (Beaver, 1997).

A linguistic context where presuppositions do not survive is represented by verbs like *say* and *tell*; they prevent the projection of the presuppositions of the sentential objects: *Bill says he is not guilty* does not presuppose he is innocent. (Karttunen, 1974) has classified these verbs as *plugs*, in order to account for their behavior. In our model, such verbs are semantically interpreted as instances of the action schema of the corresponding speech acts (see (Ardissono et al., 1998)): from the precondition of the action of informing, and under the sincerity assumption, it is possible to infer only that Bill believes the proposition he uttered, while no information is given about the speaker's beliefs. Therefore, the semantic representation consists in a mental model of the action that contains an embedded mental model representing Bill's belief that he is not guilty.

Similarly, a question performed by a speaker provides a context in which the presuppositions may be cancelled, even if there is no negation: in fact, the representation of *Did John arrive?* contains an instance of the linguistic action representing questions: since it has the precondition that the questioner does not know whether the propositional content is true, two mental models are possible: one in which John arrived and another representing *John didn't arrive*.

Finally, we want to highlight how some inferences, traditionally classified as presuppositions because of their resistance to negation and cancellability, can receive a more accurate explanation than as "preconditions for actions". (Soames, 1989) noticed the different behavior of the factive verbs *regret* and *realize*: in hypothetical contexts, the former maintains its presupposition that the speaker of the utterance believes that the content of the subordinate clause holds, while the latter does not:

If I regret that I told the false, I will confess it.
If I realize that I told the false, I will confess it.

The difference emerges when the corresponding action definitions are examined. In fact, the precondition for uttering the verb *regret* is that both the speaker and the described agent, at the event time, believe that the subordinate clause p is true.
However, on the contrary, *realize* has the precondition that p is true (according to speaker's beliefs) and that the agent who realizes does not know it is: rather, the agent's knowledge that p results from the action effect.
When these verbs are asserted in past tense, they share the presupposition that the speaker believes p: in fact, the action preconditions must be true. Moreover, if *realize* is asserted in the first person, the speaker and the described agent coincide. In the past tense, this means that only after the *realize* event happened, the speaker came to know that the event preconditions were true (i.e. p was true and he didn't know p). Were the sentence uttered in a hypothetical context, some mental models of the description would represent the *realize* event as not happened: in some models, the agent does not come to know that p has been true from the start, and that he was not aware of it, thus blocking the conclusion that in all models he is currently aware of p's truth.

Comparison with related works

Many approaches to presuppositions have a logical bias: the presupposition is interpreted as a function transforming contexts represented as set of sets of possible words (Beaver, 1997). However, as (Johnson-Laird, 1983) claims, logic is not a candidate for building cognitively plausible solutions to reasoning. In this work we have shown how mental models can be exploited to give an explicative solution to the problem of presuppositions that be also cognitively plausible.

(Marcu and Hirst, 1996) propose a treatment of pragmatic inferences which aims at accounting for both conversational implicatures and presuppositions in a single way. They introduce two different notions of satisfaction of a formula, where the first one is preferred in case of conflicts. This mechanism allows to distinguish between pragmatic inferences that can be canceled (i.e. conversational implicatures and presuppositions from negative sentences) and those that cannot be removed felicitously (presuppositions from positive sentences). For these rea-

sons, different rules referring to different satisfiability notions are needed to express the fact that the same presupposition is triggered by the same lexical item in positive and negative sentences.
Moreover, they argue that a default based formalism cannot explain pragmatic inferences because they are not always cancellable. On the contrary, we keep apart the treatment of conversational implicatures from that of presuppositions. We exploit a single nonmonotonic form of reasoning for modeling implicatures, while presuppositions are explained by the interaction of mental models with scalar implicatures. Furthermore, we need no rules for deriving presuppositions, even in positive sentences, since presuppositions emerge from the plan-based representation of action verbs.

Many approaches relate presuppositions and anaphora, exploiting the DRT formalism (Van Der Sandt, 1992; Asher and Lascarides, 1998). However, the cancellability of presuppositions is not explained on the basis of a non-monotonic framework, but is based on the notion of global vs. local accommodation of the presupposed information; i.e if a presupposition in a subordinate clause is globally accommodated it becomes a presupposition of the whole sentence.
This solution has two shortcomes: presuppositions are kept apart from the asserted content and they are first introduced in the local context of the trigger: they can be removed later if they can be accommodated in a wider scope; second, defeated presuppositions (i.e. locally accommodated ones) are still present in the local context, representing information that is not true even at the local level (consider *he didn't arrived since he didn't leave*).
As a consequence, when they face the projection problem, DRT approaches have to resort on further mechanisms like discharging the global accommodation due to the lack of the informativeness of the interpretation.

In these models, the presuppositions are triggered by lexical items in an unexplained way, instead of arising from the lexical representation. Furthermore they do not take into account the differences between positive and negative contexts, and different explanations are needed for the phenomena of implicature and presuppositions.

Finally, our solution does not incur in the problem highlighted by (Zeevat, 1992): lexically triggered presuppositions must be accommodated not only globally as in (Van Der Sandt, 1992)'s approach but also locally: in our case the presupposition is not kept apart from the verb interpretation but is related to the preconditions and effects of the action.

Conclusions

Mental models were introduced by (Johnson-Laird, 1983) as a reasoning framework that is endowed with a cognitive plausibility. Here, we tried to show how mental models can be exploited to provide natural and general explanations for linguistic phenomena.

Action verb presuppositions are ruled out as an independent phenomenon, to reappear as an opportunistic phenomenon, stemming from the interaction of many other factors. In the first place, conversational implicature, and, in the second place, the reasoning on a mental model representation. This representation, in turn, relies on a plan formalism to represent actions: mental models offer a natural way to exploit action plans, and to reason on them. Defeasibility of implicatures completes the framework, causing presuppositions to look cancelled, under certain circumstances.

References

Ardissono, L., Boella, G., and Lesmo, L. (1998). An agent architecture for NL dialog modeling. In *Proc. Second Workshop on Human-Computer Conversation*, Bellagio, Italy.

Asher, N. and Lascarides, A. (1998). The semantics and pragmatics of presupposition. *Journal of Semantics*, in press.

Beaver, D. (1997). Handbook of logic and language. In van Benthem, J. and Meulen, A. T., editors, *Presuppositions*, pages 939–1008. Elsevier, Amsterdam.

Boella, G. and Damiano, R. (1999). Plan-based event representations for the analysis of tense and aspect. Submitted to conference review.

Bratman, M., Israel, D., and Pollack, M. (1988). Plans and resource-bounded practical reasoning. *Computational Intelligence*, 4:349–355.

Gazdar, G. (1979a). *Pragmatics: Implicature, Presupposition and Logical Form.* Academic Press, New York.

Gazdar, G. (1979b). A solution to the projection problem. In Oh, C. and Dineen, D., editors, *Syntax and semantics 11: Presupposition.* Academic Press, New york.

Goy, A. and Lesmo, L. (1997). Integrating lexical semantics and pragmatics: the case of italian communication verbs. In *Proc. of Int. Workshop on Computational Semantics*, Tilburg.

Hirschberg, J. (1985). *A theory of scalar implicature.* PhD thesis, University of Pennsylvania.

Johnson-Laird, P. (1983). *Mental Models.* Cambridge University Press, Cambridge.

Karttunen, L. (1974). Presuppositions and linguistic context. *Theoretical Linguistics*, 1:181–194.

Marcu, D. and Hirst, G. (1996). A formal and computational characterization of pragmatic infelicities. In *Proc. of ECAI-96*, pages 587–591, Budapest.

Schaeken, W., Johnson-Laird, P., and d'Ydewalle, G. (1996). Tense, aspect, and temporal reasoning. *Thinking & reasoning*, 2:309–327.

Soames, S. (1989). Presupposition. In Gabbay, D. and Guenther, F., editors, *Handbook of Philosophical Logic*, pages 553–616. Reidel, Dordrecht.

Van Der Sandt, R. (1992). Presupposition projection as anaphora resolution. *Journal of Semantics*, 9:333–377.

Zeevat, H. (1992). Presupposition and accomodation in update semantics. *Journal of semantics*, 9:379–412.