Dressing Dialog for Success

Pat Healey* **& Carl Vogel** †

1 Introduction

This paper addresses two questions about dialog meaning: what is it; where does it exist? Unfortunately, we quickly realize that the first question is probably ill-directed as we don't have many pre-theoretic intuitions about how a dialog can mean something in the way that intuitions hold about even a multiply authored text. Yet, partly because (for example) jointly authored texts often do have meaning, we feel that dialog meaning also requires exploration. Further, we feel a strong intuition that an adequate characterization of a successful dialog is germane to dialog meaning, much in the same way that characterizing truth conditions express aspects of sentence meaning. However, stipulating criteria for dialog success creates a new version of the second question: where does success get decided? Ratings of dialog success hinge on the choice of perspective from which the dialog is viewed, along with that perspective-dependent determination of whether or not the interlocutors have determined that they mean different things with individual expressions or classify the world differently. Arguably, the location of meaning covaries with the choice of perspective.

Consider (1).

(1) Pat: "The cat is on the mat."

Carl: "Er. No it isn't."

It is not clear what (1) means. It is easier to consider whether (1) is successful. A plausible generalization of the usual truth conditional approach to semantics takes account of multiple agents, but is nonetheless inclined to locate them in the same world or universe; this immediately requires the annotation of predication (or some modal operation notation) of belief in order to keep the universe consistent. On this line, (1) means that Pat believes that the cat is on the mat and Carl believes that the same cat is not on the mat. We take it to be obvious that this is not adequate as an account of the meaning of the dialog — the meaning of a dialog is not given by the truth of its turns when interpreted as ascriptions of belief to agents with respect to a single ontology — for such a semantics would reduce the meaning of a dialog to the meaning of a discourse uttered by an omniscient deity.

* ATR International, Media Integration and Communications Research Laboratories 2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02, Japan, ph@mic.atr.co.jp

† O'Reilly Institute, Computational Linguistics, Trinity College, University of Dublin, Dublin 2, Ireland, vogel@cs.tcd.ie

In a Tarskian approach (like that of Runcan, 1984) , (1) is successful if Pat achieves a revised information state such that BEL(PAT,¬ON(THE CAT, THE MAT)) is true. But this is not satisfactory, because it ignores the judgements of the interlocutors. It would be a more general notion of dialog success if it did not require participants to reach the same state at the end of the exchange, permitting them instead to be in states which to an external observer are distinct but which are indiscriminable to the participants. A dialog is successful if the participants think it is.

We have been espousing a more pessimistic view about the likelihood of denotations to be shared among interlocutors than is required for the standard approach to semantics (Healey & Vogel, 1994, 1996; Healey, 1995, 1997) In this view, a sentence uttered by a speaker is true if and only if it is a token of a sentence type which adequately characterizes a state of affairs that the speaker assumes to hold (note: this is a more finely grained model of sentence meaning than one which considers sentence meaning in the absence of a speaker; moreover, there is not an explicit encoding of belief about the world as belief about the world is deemed constitutive of the world). This is more pessimistic than the Tarskian view because interlocutors are not in any strong sense assumed to inhabit the same universe. Interlocutors are weakly taken to be neighbors by virtue of the semantics of dialog being articulated solely from a third perspective, that of an observer. However, in this framework, the observer is not assumed to be God; the interlocutors themselves can act as the observers.

Let the observer's perspective be called 'encompassing', and the interlocutors' perspectives are 'local'; then, given an abstract notion of a task which functions as an intentional source for the dialog, we can articulate three possible criteria for dialog success:

1. no encompassing inconsistencies,
2. no local inconsistencies,
3. task completion:
   (a) successfully,
   (b) unsuccessfully.

The conditions in (3) are clear: the intentions accompanying a dialog may or may not be met as the result of the dialog. This is the location of the 'world' in dialog semantics, that is, as relativized to the demands placed by the task at hand on coordination. Condition (2) obtains if no interlocutor perceives there to be an interpretational conflict with respect to the other interlocutors, and (1) holds when relative to the observer's perspective there is no conflict among the interlocutors. It is fully possible for (1) to fail while (2) obtains, as well as (1) to hold when (2) fails to. A Tarskian model of dialog success would use ((1) $\wedge$ (3a)) as its criterion (with an assumed entailment (1) $\supset$ (2)). Previously we have focussed

on ((2) $\vee$ (3a)). In this paper, in part through the influence of Heydrich and Rieser (1995) , we adopt (2) as a discriminating criterion of dialog success.[1]

The paper proceeds as follows. We illustrate that the original model provided by Healey and Vogel is inadequate by providing a classification of a dialog collected during the Maze Task experiments. The original model is adjusted by incorporating a characterization of conflicting information. This allows us to provide a more finely grained analysis of semantic convergence than was possible before. With this addition it is possible to succinctly express conditions that discriminate the success of a dialog (essentially, a formal articulation of (2)). The discrimination condition offered is as optimistic with respect to dialog success as their original model is pessimistic with regard to cognizance of shared denotations — more dialogs are deemed successful by the model than a Tarskian view can countenance, but this is not a defect. Rather, we provide the minimal conditions for dialog success. Within the model it is possible to classify, for example, the success of a diplomatic dialog,[2] while, counterintuitively, a Tarskian model would be obliged to label most dialogs in this genre as failures.

2 Modeling Tools

Our model (Healey & Vogel, 1994, 1996) is developed in Channel Theory (CT) (Seligman, 1990; Barwise & Seligman, 1993, 1994; Seligman & Barwise, 1993), a mathematical framework that has grown out of the situation-theoretic analysis of constraints (Barwise & Perry, 1983; Barwise, 1989), and conditionals. CT offers a naturalized theory of information flow that encompasses both the reliability and the fallibility of natural regularities. CT is directly influenced by the work of Seligman (1990) on perspectives, as offering a means to characterize context dependent, local, informational dependencies without recourse to a globally omniscient information state.

A channel is a connection between classification domains, where a classification domain is comprised of tokens, types and the classification relation between them. Tokens are parts of the of the world which are made accessible by the types an agent classifies them with. Untyped tokens are inaccessible to the agent. A channel between classification domains includes an indicating relation between the types in the respective classifications and a signaling relation between the tokens. Indicating relations are type-level objects that model constraints. They articulate what a token that satisfies the antecedent of a constraint indicates about some other token. Signaling relations are token-level objects

[1] It is indicative of the enterprize that the authors of this paper seem to share some expressions but dispute their meanings. In the terms of this paper, we have to rate our own discussion as largely unsuccessful as we cannot agree to its text on rather substantial points. However, we won't drag this paper into a morass of disclaiming footnotes. We hereby offer a promissory note for a single manuscript in which each declares and defends his own perspective. The fact that we seem to agree about what the model is (as opposed to what it is a model of) is haunting, but in itself constitutes the modicum of successful dialog between us that keeps us hopeful and justifies our not having dustbinned the whole project.

[2] A diplomatic dialog is often designed to resolve a conflict and therefore, when successful, can contain inconsistencies at the encompassing level that are not present at any local level.

that relate the objects which actually satisfy the constraint. For example, the statement *books are expensive* can be analyzed as an indicating relation between things of type *book* and things of type *expensive*. An indicating relation is informative in situations where there is a signaling relationship between tokens of the appropriate types. *Books are expensive* is informative just if a real connection links a token of type *book* to a token of type *expensive*.[3] If an indicating relation models a constraint, then a signaling relation models a singular instance of that constraint. The capacity to deal with the connection between tokens as a real object in itself is one of the useful features of CT.[4]

It is useful to formalize some concepts:

Dfn. 1. Let $s_1, s_2, ...$ be tokens (also called sites); $c_1, c_2, ...$, connections; $T_1, T_2, ...$, types.

Dfn. 2. $s_1 \models T_1$ denotes that s_1 is of type T_1.

Dfn. 3. $T_1 \Rightarrow T_2$ denotes an *indicating relation* between types T_1 and T_2.

Dfn. 4. $c \models T_1 \Rightarrow T_2$ denotes that connection c is typed by the indicating relation $T_1 \Rightarrow T_2$.

Dfn. 5. $s_1 \xrightarrow{c_1} s_2$ denotes a *signaling relation* between tokens relative to the connection c_1; s_1 is a *signal*, and s_2 is a *target*.

Dfn. 6. A connection c is the *serial composition* of c_1 and c_2 (c_1;c_2) iff for all sites $s_1, s_2 \in S$ $s_1 \xrightarrow{c} s_2$ iff there is an intermediate site s such that $s_1 \xrightarrow{c_1} s$ and $s \xrightarrow{c_2} s_2$.

Dfn. 7. A *channel* $\mathcal{C}$ is a set of connections and their classifications $\langle C, T \times T, \models \rangle$.

Under the modeling assumption that there are no unclassified tokens, if there is a signal/target pair then both the signal and target are of some type. Their respective classifications each form a classification domain, and the connecting channel instance c is a token level connection between the classification domains. Token level connections give rise to a variety of type level properties.

Dfn. 8. A constraint $T_1 \Rightarrow T_2$ is *informative* iff $\exists c \in C : c \models T_1 \Rightarrow T_2$ and there are tokens s_1 and s_2 such that $s_1 \xrightarrow{c} s_2$ with $s_1 \models T_1$ and $s_2 \models T_2$.

Dfn. 9. A constraint $T_1 \Rightarrow T_2$ is *sound* iff $\exists c \in C : c \models T_1 \Rightarrow T_2$ and for all tokens s_1 and s_2 if $s_1 \xrightarrow{c} s_2$ and $s_1 \models T_1$ then $s_2 \models T_2$.

Dfn. 10. A constraint $T_1 \Rightarrow T_2$ has a *pseudosignal* s_1 iff $s_1 \models T_1$ but there is no s_2 or c such that $s_1 \xrightarrow{c} s_2$.

Dfn. 11. A constraint $T_1 \Rightarrow T_2$ has a *multisignal* s_1 iff $\exists c \in C : c \models T_1 \Rightarrow T_2$ and $s_1 \models T_1$ and there is more than one s_i such that $s_1 \xrightarrow{c} s_i$.

Dfn. 12. A constraint $T_1 \Rightarrow T_2$ has a *clear signal* s_1 iff $s_1 \models T_1$ and there is a unique s_i and c such that $c \models T_1 \Rightarrow T_2$ and $s_1 \xrightarrow{c} s_i$.

[3] Note that it's not necessarily the same token, but a situation which is itself classified by types corresponding to the context of an item being dear — exactly the same item in a different situation might not be expensive at all.

[4] Outlining conditions for composing channels in terms of indicating and signaling relations offers a natural way to specify a semantics for systems of defaults that depend on graph-theoretic inference procedures (Vogel, 1997).

These represent degrees of successful information flow: pseudosignal involves a failure to ground the indicated classification in a signaled token; a multisignal, the possibility of grounding the classification in many tokens; a clear signal, unique classification of a token. It is an informative constraint if it sometimes holds, and it is sound if it always holds when the antecedent is satisfied.

3 The Model

Classification domains are structures: $\langle S, T, \models \rangle$. As noted, tokens in S are accessible only via the types in T that pick them out. A classification domain determines an ontology and is relativized to an agent. Choice of classification domain conditions the possible connections among types and tokens. A **conceptualization** is a classification domain together with indicating relations and signaling relations : $\langle S, T, \models, C, \Rightarrow, \mapsto \rangle$. Intuitively, a conceptualization is intended to model the pre-theoretic notion of a perspective, with agents able to adopt more than one perspective on the same problem. We model the aspects of an agent salient to communication with a structure $\langle \Pi, \preceq, i \rangle$, where Π is the set of conceptualizations the agent works with, $\preceq$ is a preference ordering of conceptualizations, and i is an index to the agent's current working perspective.

3.1 Intra-Agent Structure

The first part of the analysis is a more structured model of sentence meaning for an individual agent. Later we address the implications this model has for the location of meaning. Consider (2).

(2)
$$\begin{array}{ccc} \langle\langle ON, cat, mat; 1\rangle\rangle & \Longrightarrow & \left[\begin{array}{l}\text{phon: } \textit{the cat is on the mat} \\ \text{synsem: } [\text{ content: p}]\end{array}\right] \\ \amalg & & \amalg \\ s^1 & \xrightarrow{c} & s^2 \end{array}$$

We assume that there is a constraint between idea types and sentence types that is strongly tied to instances of ideas and individual utterances. It has long been understood that the import of an individual use of a sentence is distinct from the meaning of the sentence divorced from a specific occasion of use. Here we pay attention to both aspects. There is some token (s^1), a part of the world (say, Pat's brain state in some situation), which is classifiable in terms of the type obtainable from the infon $\langle\langle ON, cat, mat; 1\rangle\rangle$, such a classification is taken to represent the individual having the idea so typed. Ideas are systematically related to utterance types. A particular constraint relating an idea to an utterance is informative just if that constraint is the type of a connection from some particular instance of uttering to some particular token of having the idea. The constraint is a type-level object that serves as the classification of tokens: the connection between an utterance token and an idea token. For an individual generating an utterance, such a

connection has to be in place and of the right type for it to be possible to classify there having been an utterance expressing the idea.

The constraint between idea types and utterance types needn't be universal: there are any number of ways of expressing an idea. We assume that sentences are also modeled by types, and we use a simplified HPSG (Pollard & Sag, 1994) notation to describe the linguistic objects that classify individual utterances. The significance of this theory for our model is that the linguistic types include information about both syntax and semantics (and aspects of pragmatics, but that is not relevant at the moment). Thus, our model of sentence meaning includes (in addition to utterances of the sentence, and syntactic information available from utterance types) two sorts of semantic content: linguistic content (target content) and signaling content. In (2) we have expressed the linguistic content as just p, as a placeholder for the *linguistic* representation of the utterance meaning. We refer here to the information about semantics that is encoded in the utterance type, the value of the CONTENT attribute in the sign, which is a structured cue to what meaning is. This is distinct from the *meaning* of the utterance type, but is obviously related because of an informative constraint.[5] The relationship between the signaling content and the target content is important to the notion of ambiguity. Those who assert that for speakers, sentences are unambiguous (Burton-Roberts, 1994) are really asserting something about the signaling content rather than the target content. The information in p may indeed be ambiguous (as when there is quantifier scope ambiguity) when the idea that signals an utterance type that includes p as its content is not.

Information flows from utterance to interpretation as well; (2) depicts the content of a production channel. An interpretation channel involves constraints in the opposite direction. It turns out to be most efficient under idealized dialog conditions and evolutionary models to use something (equivalent to) bidirectional constraints rather than completely distinct channels for production and interpretation (Hurford, 1989), thus yielding something like a Saussurean sign. The channel in (2) is informative for the agent; i.e., there is a signaling relation $s_i \xrightarrow{c} s_u$ and $s_i \models T_i$ and $s_u \models T_u$. Further, we admit faulty conceptualizations. For example, there may be pseudosignals. Although some token has been correctly classified as an instance of some concept type antecedent in a constraint, it does not signal an utterance of the indicated type, as illustrated in (3) and (4).

(3)
$$\begin{array}{c} \langle\langle ON, cat, mat; 1\rangle\rangle \\ \bigsqcup \\ s^1 \end{array} \implies \left[\begin{array}{l} \text{phon: } \textit{the cat is on the mat} \\ \text{synsem: } [\text{ content: p}] \end{array}\right]$$

[5] Another aspect of this distinction is that while we presume evolution to have selected a roughly similar ontology of features in a linguistic type, we make no such assumption about the internal structure, if any, of idea types.

$$(4)\qquad \begin{array}{c} \langle\langle ON, cat, mat; 1\rangle\rangle \\ \Vdots \\ s^1 \end{array} \quad \Longrightarrow \quad \begin{array}{c} \left[\begin{array}{l} \text{phon: } \textit{the cat is on the mat} \\ \text{synsem: } \left[\text{ content: p}\right] \end{array}\right] \\ \Vdots \\ s^2 \end{array}$$

These two possibilities are both pseudosignals, but of different sorts. Consider them as indicative of the state of an entire conceptualization, with respect to these types (i.e., no other tokens supporting the types instead): in (3) there is no perspective in the conceptualization which has an utterance token classified by the linguistic type — the agent has the idea, and a constraint[6] stipulating that some utterance type expresses the idea but under no circumstances would the agent use that expression in an utterance (perhaps seems far-fetched at first blush, but this does serve as an analysis of the *tetragram*); in (4) (also under the hypothesis that this is true of all of the agents' perspectives) we have a circumstance in which it is conceivable that the speaker would use the utterance type in connection with the idea, but no actual use — no actual connection between utterance token and idea token related to this constraint (but, perhaps related to some other constraint, for instance in quotation). These notions are also available within a single perspective: a pseudosignal in the agent's preferred perspective might be a clear signal or a multisignal in another.[7] In terms of production, a multisignal arises as the result of synonymy, or underspecified conversational demands.[8] These conditions on an individual's conceptualization are summarized in Table 1.

This concludes the first part of the model of dialog meaning; it is in terms of utterance meaning for an individual. The meaning of an utterance (sentences included) is a relation between a rich linguistic type, a concept type (which needn't be linguistic), tokens that are classified by those types, and a further classification of the connection between the tokens in terms of the relations between the types. In essence, under just this much of the model, meanings are, with some qualification, in the head. This follows from our assumption that no tokens are unclassified: by classifying the world conceptually and linguistically, the individual creates both meaning and the world. However, the caveat is that we presume there to be a world that is being classified. This means that this cannot really count as 'narrow' construal; however, we still take it as the version of meanings-being-in-the-head that is appropriate to an Austinian notion of propositions. We also assume that certain classifications may prove to be more optimal than others. In particular, we take it as a fact of evolution and biology that the complex linguistic types that classify utterance tokens are closely aligned across individuals in a community. We presume that evolution and biology select for a similar ontology in the structure of linguistic types, even though we don't assume that any individuals have the same linguistic types even up to isomorphism. Nonetheless, this means that the types are distinct. It remains to characterize how a dialect

[6] An uninformative one.

[7] If it's a pseudosignal in just the preferred perspective, then it's a *weak* pseudosignal; it's a *strong* pseudosignal if it's a pseudosignal over all perspectives in the conceptualization.

[8] A multisignal is a *strong* multisignal if it is a multisignal in the preferred perspective; it is a *weak* one if it arises from alternative classifications in other perspectives within a conceptualization.

can emerge from a set of idiolects, and how both the public and private aspects of meaning are accommodated within the model.

Signal Type	Production C	Interpretation C^{-1}
Clearsignal	Concept type indicates a unique utterance type under the current conceptualization	Utterance type indicates a unique concept type under the current conceptualization
Pseudosignal	Concept type does not indicate an utterance classified by any type under the current conceptualization (e.g., inarticulable or ineffable)	Utterance type does not indicate a concept classified by any type under the current conceptualization (e.g., slip of tongue)
Multisignal	Concept indicates more than one utterance type. (e.g., vague formulation)	Concept type is indicated by more than one 'utterance' type (e.g., multi-modal communication)

Tab. 1: Some Possible Characterizations of Production and Interpretation

3.2 Inter-Agent Structure

Building on the model of an agent in terms of an individual conceptualization, we model communication via interacting individuals, under the assumption that communicating agents do not have the same set of conceptualizations as each other or third-party listeners. Only an omniscient agent with a "god's eye view" could ascertain communication among fallible agents whose types and tokens are potentially disjoint. Figure 1 illustrates this condition. The theorist/observer is an essential part of the model of communication, and therefore meaning in dialog. Obviously communication does not require the presence of a third distinct party, however the role of observer is nonetheless necessary. When there actually is no third party involved the picture collapses into an account of mutual modeling, albeit one in which an agent is assumed to model a partner in terms of the agent's own conceptualizations and not the partner's. We emphasize that it is not a prerequisite to successful communication that participants have intersecting ontologies. The only assumption we do make is the barest minimum: communication occurs just in case to some observer (who may be coextensive with one of the participants), there is some signaling relation which has been assigned a type by both interlocutors. That is, communication is seen as the articulation and evolution of channels between agents, yet this channel is that of an observer — either a participant/observer or a wholly external one. The token act of communication is modeled as a signaling relation, the content of which is determined by its type, this classification itself 'owned' by the observer.

A listener, the theorist for instance, has her own ontology which includes both P_A and

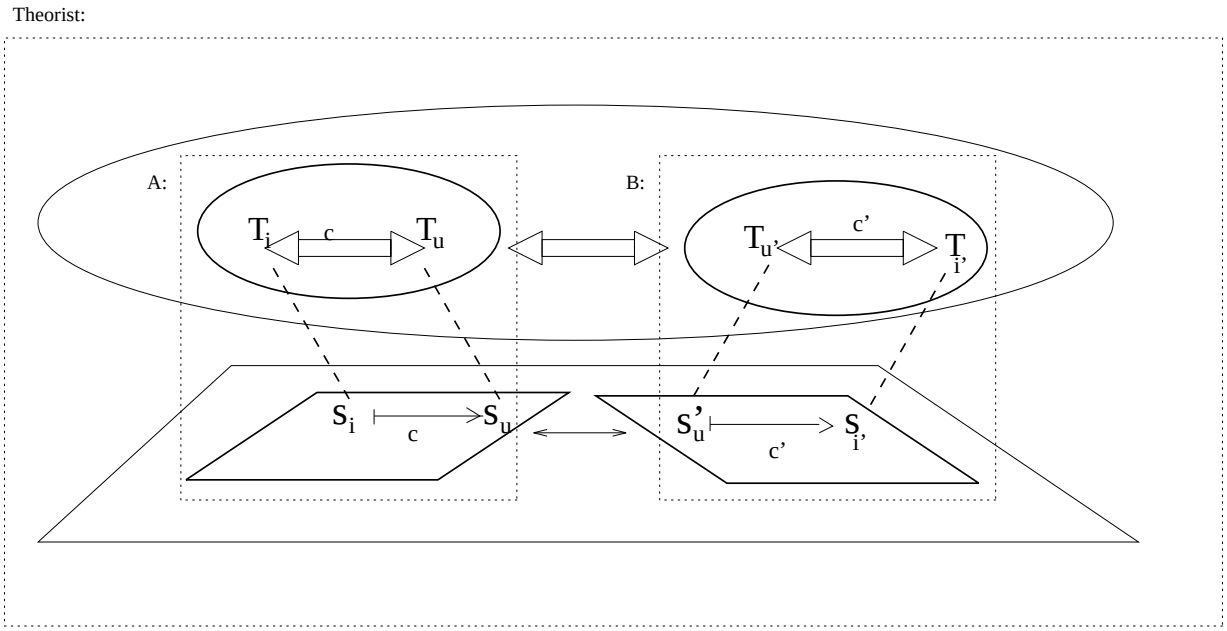


Fig. 1: Inter-Agent Signal

P_B as tokens, as well as a classification of the types and tokens that the listener determines each of the interlocutors to possess (the listener does not own the same types and tokens as the interlocutors except in the special case of omniscience). A communication act is a token-level object owned by the listener which forms a connection between the tokens 'theoretically owned' by the interlocutors and the connection is typed by a constraint on the respective utterance types.

Dfn. 13. Given two dialog participants $\mathrm{P}_A = \langle \Pi_A, \preceq_A, p_A \rangle$ and $\mathrm{P}_B = \langle \Pi_B, \preceq_B, p_B \rangle$, an *observer* is a distinguished agent $\mathrm{P}_O = \langle \Pi_O, \preceq_O, p_O \rangle$ where $\Pi_O \supseteq \Pi_O^A \cup \Pi_O^B$ and similarly for $\preceq_O$ and p_O such that $\mathcal{R}_\Pi(\Pi_O^A, \Pi_A)$ and $\mathcal{R}_\Pi(\Pi_O^B, \Pi_B)$ (similarly, $\preceq$ and p) where $\mathcal{R}$ is a mapping between the two structures. Let $\mathrm{P}_{O_A} = \langle \Pi_O^A, \preceq_O^A, p_O^A \rangle$ and let $\mathrm{P}_{O_B} = \langle \Pi_O^B, \preceq_O^B, p_O^B \rangle$.

Dfn. 14. For an observer P_O, if $\mathcal{R}_\Pi$, $\mathcal{R}_\preceq$ and $\mathcal{R}_p$ are isomorphisms, then the observer is *God*.

We now define a few more notions important to the model of dialog meaning.[9] With the exception of (16), these definitions are those presented by Healey and Vogel (1996).

Dfn. 15. Let P_O be an observer, $Cl(\mathrm{P}_O)$ is the agent with signaling and indicating relations closed under serial composition (an *attentive observer*).

Dfn. 16. A *dialog* is a sequence of attentive observers:

$\Delta = \langle \langle Cl(\mathrm{P}_{O_1^1}), \ldots, Cl(\mathrm{P}_{O_1^i}) \rangle, \ldots, \langle Cl(\mathrm{P}_{O_n^1}), \ldots, Cl(\mathrm{P}_{O_n^i}) \rangle \rangle, 1 \leq n,$

[9] Of course, all generalize in obvious ways for instances of n-ary communication as well.

where n is the number of successive communication acts, and i is the number of alternating observers.

Dfn. 17. A *turn* is a dialog in which $|\Delta| = 1$.

Dfn. 18. A participant P_A has a *signal* s_1 *for a participant* P_B according to an observer P_O if s_1 is a site in P_{O_A} and there exists a channel $c \in C$ in $Cl(\mathrm{P}_O)$ such that $s_1 \xrightarrow{c} s_2$ for some site s_2 in P_{O_B}. We also say that P_A has a signal for P_B *through* c', where either c' is c, or c is a composite channel with c' as one of its components.

In Dfns. 15 and 18 the composition referred to is just the serial composition of channels defined in Section 2. Dfn. 16 formalizes communication in dialog from the perspective of a set of attentive observers as the succession of 'mental states' induced by the corresponding succession of communicative acts observed.

One interlocutor has a signal for another if the speaker's idea is connected (according to the observer's vantage point) to an utterance situation which the addressee also classifies. A signal defines a structural condition for information flow between the agents. However, information can flow between agents without both agents ending up with the same piece of information, precisely because the agents need not share any types or other tokens. The speaker's information gives rise to an utterance which, if classified by the addressee, is understood in the addressee's own terms. The most ideal form of communication is when the signal is clear: interpretable and unambiguous in the hearer's preferred working perspective.[10]

Dfn. 19. An agent $\mathrm{P}_A = \langle \Pi_A, \preceq_A, p_A \rangle$ has a *clear signal* for $\mathrm{P}_B = \langle \Pi_B, \preceq_B, p_B \rangle$ if P_{O_A} has a site s_1, P_{O_B} has a site s_2, there is a channel $c \in C$ such that $s_1 \xrightarrow{c} s_2$, and s_1 is neither a pseudosignal nor a strong multisignal for P_B (i.e., s_2 is an internal site and c is unique when restricted to channels from P_{O_B}'s current perspective).

A clear signal models 'successful' communication. Crucially, when a speaker has a clear signal for an addressee, there is no requirement that both interlocutors are thinking of the *same* thing. All it implies is that their interpretations are *mutually indiscriminable* with respect to the current state of the dialog. It is entirely consistent with this idea that it may transpire during the course of the dialog that the interlocutors had adopted different interpretations. In this model, "talking about the same thing" is contingent on the goals of the dialog; only an omniscient observer could determine whether agents are *really* talking about the same thing.

Table 2 provides a summary of some of the discriminations that are possible for communication and miscommunication in dialog. When interlocutors detect the conditions specified on the left hand side of the table they are likely to respond in the ways suggested on the right.

Note that even signalhood is defined relative to the observer's perspective. An utterance is interpreted in the addressee's own conceptualization as it exists to the observer. Picture

[10] The notions of pseudosignal and multisignal are also exactly those from Healey and Vogel (1996) , they just generalize from one agent to more than one via intermediate sites owned by the observer.

Condition	**Reaction**
Clear Signal	Move to Next Turn
Internal Pseudosignal	Self Repair
Weak Pseudosignal	Uninterpretable in Current Perspective
Strong Pseudosignal	Uninterpretable in any Perspective
Strong Multisignal (low degree)	Specific Clarification
Strong Multisignal (high degree)	General Clarification
Weak Multisignal	Philosophizing

Tab. 2: Semantic Discriminations in Multi-Agent Communication

a dialog between two people with no actual third party observer. Communication happens because the role of observer shifts between the two participants. Suppose P_A makes an utterance. When we take P_A as the observer, then presumably the utterance relates to a clear signal. When we take P_B as the observer of the same exchange, there is perhaps a pseudosignal for P_B. The dialog is modeled as the alternating sequence of attentive observers. So, there is 'clear' communication throughout the dialog when there is an isomorphism of conceptualizations of observers throughout that sequence. There are three levels of structure that are interesting with respect to isomorphisms across individuals: internal feature ontologies of types, classification with types and constraints of tokens and connections by interlocutors, classification with types and constraints of tokens and connections by observers. We are not addressing isomorphisms of the first sort. We also doubt that there is ever one of the second sort — we defined God, who would be able to determine if the second sort existed, in terms of an isomorphism between pairs of interlocutors and the observer. It is important to emphasize that this isn't about the internal structure of the types, only the arrangement of types with respect to tokens. Clear communication is defined by isomorphisms in these classifications among the observers. This is a more complicated picture than sketched in the introduction — there we talked about the observer in the generic; in this more complicated presentation the 'encompassing' perspective is that fixed by the entire sequence of observers (who may be dialog external), while the local observational perspective is a single observer/participant.

Meaning is here relativized to the alternation of attentive observers. There is private meaning for the individual, and public meaning to the observer. Yet, because the alternating sequence of observers do not have privileged access to information owned by the observed, the actual publicness of meaning is precisely in the isomorphisms in classification of speech acts: failing to discern differences in meaning does not impede the interlocutors from thinking that meaning has been shared, each as individual observers, and the same holds at the level of truly external observers unless they're omniscient. Electing an 'expert' to adjudicate meanings is selecting a particular observer whose classifications are deferred to. *Qua* theorists, we tend to act as if the number of alternating observers is just one, and analyze conversational data as if we know what each interlocutor means and refers to. The varying conflicting notions of where meanings exist follow from equivocating about

the number of alternating observers, election of particular observers as determining, and degree of ontological transparency of interlocutors to observers. There are public aspects of meaning (which are available only when one assumes that an observer interacting in the world is important; i.e., when one concludes that there is more than just a solipsistic agent), and there are private aspects of meaning (as obviously there are idiolects, and equally obviously idiolects are ontologically opaque to cohort interlocutors). That is, we feel that we have provided a model in which a number of conflicting senses of *meaning* can be coherently located. Furthermore, we hope to have clarified that "Where do meanings exist?" is an ill-put question, as it presupposes that *meaning* is an (perhaps the only) expression in the language with a single sense.

4 A Dialog

In this section, we demonstrate how the model captures aspects of meaning in human dialogs. The excerpt comes from a corpus generated by experiments on task-oriented dialogs (see Healey (1995) for a more detailed account). In outline, the task consisted of individuals, working in pairs, on a paper version of the Maze Task (Garrod & Anderson, 1987). Members of each pair negotiate descriptions of target positions, marked by a bold circle, in a maze-like grid (see Figure 2). For each item, one member of the pair has a picture of the maze with the target marked, the other has a picture of exactly the same maze configuration but without the target marked. The aim being to indicate, using a pen, where on the unmarked grid they believed the target location to be. The task was performed with participants seated opposite each other at a desk but with a low partition between them that preserved eye contact while preventing them from seeing each other's maze.

Depending on each pair's progress there were up to fourteen items that could be completed on each trial. The sample exchange concerns the first item in the second trial. The pair was made up of two females aged 30 and 28 who were familiar with each other.[11]

A:$_1$ right it's at the top of your page and it's the third box from your right,
B:$_2$ third box from my right?
A:$_3$ oh well it's my right
B:$_4$ your right so that would be one two three, ehh the proper box?
A:$_5$ eh?
B:$_6$ the proper box? the not the unbroken box you're talking about unbroken the third unbroken box,
A:$_7$ i don't know what you mean by third unbroken box,
B:$_8$ third box at the top?
A:$_9$ at the top aye
B:$_{10}$ right okay,

[11] Notational conventions: "[n]" indicates the point at which item n was completed in a trial. "[utterance]" in adjacent turns indicates overlapping talk. "," indicates a short pause, "..." indicates a long pause

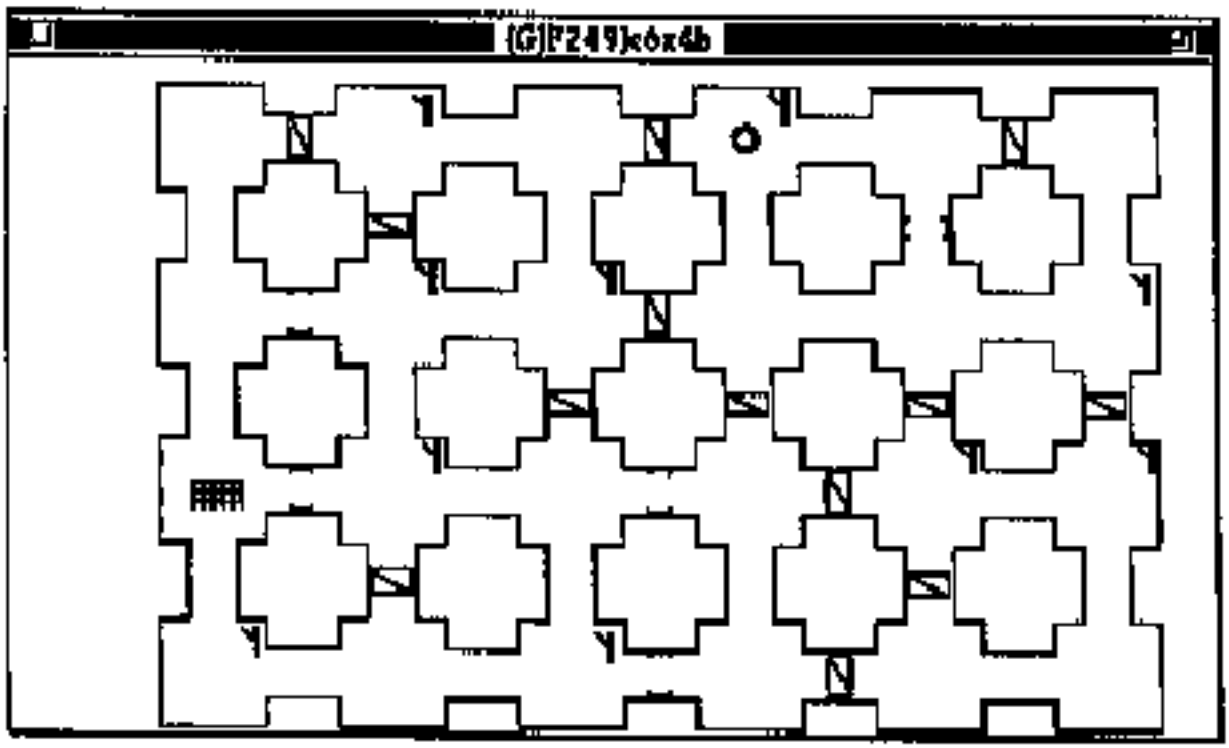


Fig. 2: The circle marks the location P_A is talking about.

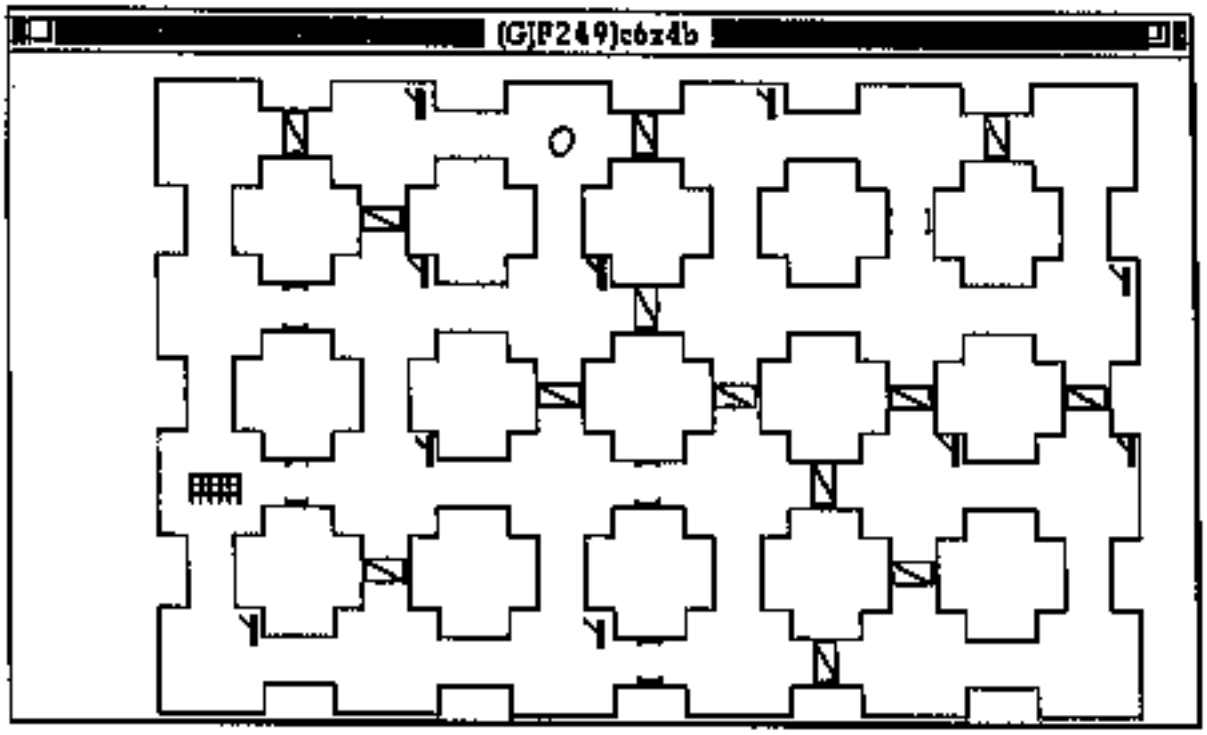


Fig. 3: The circle marks the location P_B understands P_A to have indicated.

B:$_{11}$ right right now
A:$_{12}$ oh no it's not unbroken i see what you mean no [it's] a broken box
B:$_{13}$ [yeah]
B:$_{14}$ it's a broken box
A:$_{15}$ it's very top right?
B:$_{16}$ yeah i've got it on a broken one yeah, [1]
A:$_{17}$ now

This dialog is useful in illustrating certain features of our model. Consider turn 4. An idea equivalent to $[\![$*the third proper box from the right?* $]\!]$ is linguistically realizable in a counting expression: *one, two, three, ehh the proper box?* That is, there is a content to the idea quite distinct from the semantic content of the utterance type actually employed. This is an example of what we mean to discriminate in constraints between idea types and utterance types. Further, there is a weak pseudosignal which we perceive as observers, and which P_A presumably perceives *qua* observer, in that P_B's use of the expression *the proper box* is initially uninterpretable to P_A (hence the *eh* in turn 5). However, by turn 12, P_A is able to shift to a different working perspective. Nonetheless, at the end of the dialog, although both participants seem content in thinking they are referring to the same

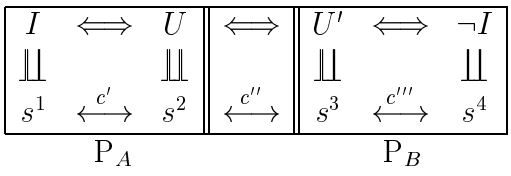
location in the maze, we (as observers) can see that they had incompatible reference. Their dialog was locally consistent, and therefore successful. Acting as individual observers they perceive themselves as meaning the same thing. Putnam would insist that they *meant* different things. Acting as privileged observers, we too perceive them as meaning different things. Some of us would argue that Putnum writes as if *meaning* is that which the privileged observer determines, however it is also possible to coherently argue that meaning is determined at the local perspective. Regardless of where meaning is determined, we take dialog success to be determined by the individuals involved. If they perceive no inconsistencies and conclude that it has been a successful dialog, then it has been, even if it fails the corporeal task at hand, as in this example, or as found by Heydrich and Rieser (1995).

At this point, we need to make clear that we have not yet detailed what an inconsistency amounts to in this model, be it local or global, and thus cannot yet describe phenomena like the fabricated dialog in (1), although we are able to describe other forms of 'miscommunications' as above. A pseudosignal is not a direct model of what we mean by inconsistency, which we take from classical systems to involve directly conflicting types rather than the absence of an object supporting a type as is the essence of a pseudosignal. We define inconsistency in terms of antisignals as in (20).

> **Dfn.** 20. An agent $\mathrm{P}_A = \langle \Pi_A, \preceq_A, p_A \rangle$ has an *antisignal* for $\mathrm{P}_B = \langle \Pi_B, \preceq_B, p_B \rangle$ iff P_{O_A} has a site s_1 which is a weak pseudosignal for P_{O_B}, with a constraint $U \Leftrightarrow I$, and all tokens t in the working perspective such that $t \models I \vee \neg I$ are also such that $t \models \neg I$.

A turn involves a local inconsistency if there is an antisignal for one of the participants through one of the attentive observers in the turn and that observer is also one of the participants. A turn involves a global inconsistency if it is inconsistent for all of its observers. In terms of information flow, an inconsistency involves a pseudosignal with respect to the channel used in the communicative act, and a clear signal with respect to one not actually used but involving a conflicting type. The notion of global inconsistency is distinct from a privileged observer's being able to say that the interlocutors mean different things by their perceptions of the utterance types at stake (the same state could constitute a clear signal to a non-privileged observer). In (5),[12] only a privileged observer can see that distinct and incompatible types are assigned to the ideas. A non-privileged observer will perceive a clear signal, unless there emerges conflicting information elsewhere (such as analysis of task failure). That is, by using the term 'globally inconsistent' we do not defer to a privileged perspective.

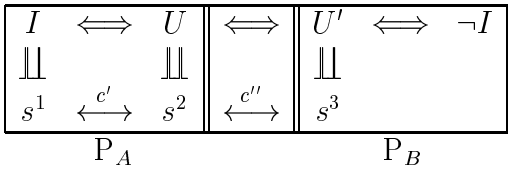
(5)



[12] This is essentially just Figure 1 in but with a slightly different configuration of types.

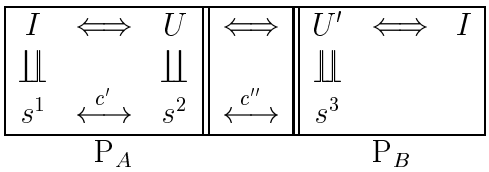
Rather, a local inconsistency involves something like (6),

(6)



in conjunction with notice that all salient[13] tokens s^i are such that $s^i \models I$. In this example, there is a local inconsistency even though a privileged observer would disagree, determining that the participants actually have the same idea I (and explaining that $\neg I$ is indicated follows from P_B misclassifying s^3 as U'). The local inconsistency follows from the fact that within the non-privileged perspective there is a pseudosignal and support for information that contradicts what would have been conveyed if the pseudosignal were clear. As another example, one in which a privileged external observer could say that they actually meant the same thing by the utterance type, is given in (7), where all all salient tokens s^i for P_B are such that $s^i \models \neg I$. This also involves a local inconsistency for P_B as observer.

(7)



Local inconsistency is an inconsistency for an observer/participant. A global inconsistency is one that is apparent to all observers. A dialog is successful if it does not contain a local inconsistency.

5 Discussion

We would argue that this is precisely the notion of dialog that is at work in diplomatic negotiation. The dialog may be quite unsuccessful in achieving what is (or what a sequence of observers might think really qualifies as being) peace, yet as long as the parties leave the table without directly conflicting on terms, the dialog is successful. Note that 'not directly conflicting on terms' doesn't presume that they mean the same thing, so long as there is an isomorphism among observers' classifications.

This seems to be an adequate notion of dialog success because it captures conflicting perspectives on what counts as success. Given pessimism about actually shared ontology, we recover optimism by assuming that the dialog is successful only if no one notices that the final turn hasn't been. Noticing lack of success is just finding the indicated type from a weak pseudosignal as inconsistent with the types that classify salient tokens in a preferred perspective. A stronger notion of (lack of) success would require there to be a local inconsistency for all observers of a turn. Still stronger is to appeal to a privileged

[13] Those that support I or $\neg I$ at all in the current perspective.

observer with access to the participants' ontologies and commensurability relations among them, who can adjudicate when conflict exists or not. Our favorite notion of dialog success is optimistic because it is possible to discriminate between the individuals — one who perhaps thinks it has been a success, the other not — and the observational perspectives. We can label the model with the same variances in the location of meaning.

We hasten to point out that this is a different interpretation than that supplied by Healey (1995, 1997) .[14] The difference is that Healey (1997) (for example), accepts the essentialist thrust of Putnam's (1975) arguments against a narrow construal of content (and the accompanying claim that meanings are in the head). Healey's solution is an attempt at naturalizing meanings. Whereas here we explore the relation between meaning for the individuals and meaning for observers, Healey considers the regularities among interactions between individuals to be prior in the explication of meaning.

References

Barwise, J. (1989). Conditionals and conditional information. In *The Situation in Logic*, pp. 97–135. Centre for the Study of Language and Information, Stanford University, CA. CSLI Lecture Notes Number 14; originally appeared in Traugott, et al. (eds), *On Conditionals*, Cambridge University Press (1986).

Barwise, J. & Perry, J. (1983). *Situations and Attitudes*. Cambridge, MA: MIT Press.

Barwise, J. & Seligman, J. (1993). Imperfect information flow. In *Proceedings of the Eighth IEEE Symposium on Logic in Computer Science*. Washington, D.C., IEEE Computer Society.

Barwise, J. & Seligman, J. (1994). The rights and wrongs of natural regularity. In Tomberlin, J. (Ed.), *Philosophical Perspectives, Vol 8*, pp. 331–65. California: Ridgeview.

Burton-Roberts, N. (1994). Ambiguity, sentence and utterance: a representational approach. *Transactions of the Philological Society*, *92*(2), 179–212.

Garrod, S. C. & Anderson, A. (1987). Saying what you mean in dialogue: a study in conceptual and semantic co-ordination. *Cognition*, *27*, 181–218.

Healey, P. (1995). *Communication as a Special Case of Misunderstanding: Semantic Coordination in Dialogue*. Ph.D. thesis, Centre for Cognitive Science, University of Edinburgh.

Healey, P. & Vogel, C. (1996). A semantic framework for computer supported cooperative work. In Connolly, J. & Pemberton, L. (Eds.), *Linguistic Concepts and Methods in CSCW*, pp. 91–107. Berlin: Springer Verlag.

[14] We further emphasize that this paper should be construed as Healey's rejecting the other interpretation.

Healey, P. (1997). The public and the private: two domains of analysis for semantic theory. In Ginzburg, J., Khasidashvili, Z., Vogel, C., Levy, J.-J., & Vallduvi, E. (Eds.), *The Tbilisi Symposium on Logic, Language and Computation: Selected Papers*, chap. 5, pp. 59–72. CSLI Publications.

Healey, P. & Vogel, C. (1994). A situation theoretic model of dialogue. In Jokinen, K. (Ed.), *Pragmatics in Dialogue Management*, Vol. 71, pp. 61–79. Gothenburg Papers in Theoretical Linguistics. Proceedings of the Special Session within the XIVth Scandanavian Conference of Linguistics and VIIIth Conference of Nordic Linguistics. August 16-21, 1993.

Heydrich, W. & Rieser, H. (1995). Public information and mutual error. Tech. rep. Report 95/11, University of Bielefeld. Situierte Künstliche Kommunikatoren, SFB 360.

Hurford, J. (1989). Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua*, *77*, 187–222.

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

Putnam, H. (1975). The meaning of meaning. In Gunderson, K. (Ed.), *Language, Mind, and Knowledge, Minnesota Studies in the Philosophy of Science, 7*. University of Minnesota Press.

Runcan, A. (1984). Towards a logical model of dialogue. In Vaina, L. & Hintikka, J. (Eds.), *Cognitive Constraints on Communication: Representations and Processes*, pp. 251–66. D. Reidel, Dordrecht, Holland.

Seligman, J. (1990). Perspectives in situation theory. In Cooper, R., Mukai, K., & Perry, J. (Eds.), *Situation Theory and its Applications, Volume I*, pp. 147–191. Centre for the Study of Language and Information, Stanford University, CA.

Seligman, J. & Barwise, J. (1993). Channel theory: toward a mathematics of imperfect information flow. Unpublished manuscript.

Vogel, C. (1997). A generalizable semantics for an inheritance reasoner. In Ginzburg, J., Khasidashvili, Z., Vogel, C., Levy, J.-J., & Vallduvi, E. (Eds.), *The Tbilisi Symposium on Logic, Language and Computation: Selected Papers*, chap. 8, pp. 103–121. CSLI Publications.