MunDial'97

Munich Workshop on
Formal Semantics and Pragmatics of Dialogue

University of Munich

March 10-12, 1997

Proceedings

edited by
Anton Benz & Gerhard Jäger

Contents

Communication Modelling and Context-dependent Interpretation:

An Integrated Approach*

René Ahn† **& Tijn Borghuis**‡

1 Introduction

Language solves a problem. It helps people to exchange ideas, even if these people come from different backgrounds, know different concepts and individuals, and have wildly diverging views in many different matters. ¿From a computational point of view, it is not obvious how language can be used to accomplish such a complicated task.

In this paper we present a simple model of communication, hoping to bring some light into these matters. As in (Ahn 1994), we assume that communication takes place between two agents, where each agent has a private and subjective knowledge state, and the knowledge of both agents is partial, finite, and represented in a computational way. In this model we investigate how ideas can be transferred from one agent to the other one, in spite of the subjective nature of the knowledge of both participants. If we pose the problem in this way, mechanisms for context-dependent interpretation seem to be a prerequisite for succesfull communication.

Our aim is not to set up a descriptive theory of language use, but to construct an explanatory model that may help to hypothesise why different language features are needed, given the problem that language helps to solve. Another feature of our model, which follows from its computational nature, is that it can be implemented on a computer. In the future this may allow us to experiment with the model in various interesting situations, for instance, cases were a misunderstanding arises between the two agents.

1.1 Ideas and Reality

The physical environment is a huge and complex structure which is accessible through the senses, and which forms the source of an ongoing stream of sensorial impressions. However, a human that interacts with such an environment, does not experience these impressions

* This research was sponsored by the Organization for Inter-university Cooperation between the universities of Tilburg and Eindhoven (SOBU).

† Tilburg University, P.O.Box 90153, 5000 LE Tilburg, The Netherlands. E-mail: ahn@kub.nl

‡ Eindhoven University of Technology, P.O.Box 513, 5600 MB Eindhoven, The Netherlands. E-mail: tijn@win.tue.nl

as a chaotic 'sensory soup' but as an organised whole, in which various familiar phenomena interact in more or less predictable ways. This is possible because the various sense impressions exhibit some redundancy, regularity, and even periodicity. The various correlations between sense impressions allow concepts to be formed that are subsequently used to 'understand' the diverse experiences from which they have arisen. Some of these concepts correspond to 'classes' or 'sets', and are 'inhabitable' , i.e. they may have instances. The person which is familiar with a specific inhabitable concept will recognise an instance of this concept, whenever he runs into it[1]. In this way he is able to connect his raw experience with the subjective concepts that he uses to classify it. Having recognised instances of various concepts within a specific situation, he can predict various things about them.

Which concepts a person has formed at a given moment in time is obviously dependend on many factors, like the physical and cultural environment that he is accustomed to, his personal history, and his specific sensorial abilities. In our model we are not concerned with he formation of concepts, and we will simply assume that each agent has his own set of concepts, often similar to but not necessarily identical with those of his fellow-agents.

1.2 Knowledge

Each person understands the world in terms of his own concepts. His conscious knowledge[2] about the world must be formulated entirely in terms of the concepts that he recognises. This knowledge is not static, but can grow as a result of communication, observation and inference processes. The resulting body of knowledge will not be a bare set of facts, but a structured conglomerate of justified beliefs, where each new item must be embedded in the knowledge which is already present. Thus, this body of knowledge is:

> *Subjective*: It is formulated in terms of personal concepts, it will be partial, and it may even be incorrect.
>
> *Incremental*: The ways in which this body of knowledge can be extended depend on what is already present.
>
> *Justified*: Knowledge is not a collection of bare facts, but will be justified in terms of other, more basic knowledge.

1.3 Modelling Knowledge States

How can the 'subjective', 'incremental', and 'justified' knowledge states that we have been discussing earlier, be formalised? Fortunately, a similar problem arises in mathematics. The main concern of mathematicians is to show which consequences follow from certain

[1] How this happens is irrelevant here, an obvious possibility is through neural networks.

[2] Here we take knowledge in the everyday sense of the word. We quote the definition in Webster's new dictionary of synonyms (p. 481): 'Knowledge applies not only to a body of facts gathered by study, investigation, or experience but also to a body of ideas acquired by inference from such facts or accepted on good grounds as truths.'

assumptions. On the one hand this activity is virtually unconstrained: new concepts may be developed, and assumptions can be freely made, independent from external reality. On the other hand the mathematician has to adhere to a strong kind of mental hygiene: concepts can only be formed if they fit into existing categories, assumptions can only be made if they are meaningfull in the context of that which is already given, and all conclusions have to be thoroughly justified.

Although this mental activity is independent from any reference to an external world, it is nonetheless quite rigourous, and can be formalised. Pure Type Systems (PTSs, Barendregt 1992) are typed lambda-calculi that can be used to record such mathematical activity in a formal and machine-readable format. They offer much freedom of expression, whilst enforing a rigorous standard of mathematical correctness.

In this paper, we assume that a person that tries to understand the outside world is, in many respects, comparable to a mathematician. However, the concepts that this person develops will be inspired by his sense data, and the assumptions that he makes are assumptions about the outside world, and are (one hopes) supported by what he sees. In other words, the whole 'body of hypotheses' that this person develops will be grounded in the external world. Taking this difference into account, knowledge states can be represented within the type-theoretical framework, and we can construct a simple model of such a knowledge state (which incorporates the participant's ability to observe (Sect. 2.3) and to reason (Sect. 2.4)).

1.4 Modelling communication

Our model forces us to adopt the view that language relates to *ideas* about reality, and not to reality itself. Communication is a process in which ideas that are privately known to one agent, is transformed in a public message and subsequently decoded by another agent, who interpretes this message, and reacts on it. If we want to model this proces starting from subjective knowledge states, we have to show how content which is meaningfull in the subjective knowledge state of one agent can be transferred to the subjective knowledge state of another agent by means of a common language.

Throughout this paper we concentrate on the simple case where two agents communicate about their common (physical) environment. Specifically, we use examples in which two agents discuss an electron-microscope, a situation taken from the 'DenK-project' (Ahn et al. 1994). In this project we construct a man-machine interface which reflects the approach to communication sketched in this paper.

2 Formalizing Knowledge states

In this section, we formalise the knowledge state of agents in the framework of PTSs. After a short introduction to PTSs, we proceed by showing how a 'context' in a PTS can be used to model a subjective knowledge state, which is linked to an external reality through the notion of 'interpretations'. To describe the effects of reasoning, observation, and communication on the knowledge states of the agents, we give an account of how these states can 'grow' as new information becomes available.

2.1 Pure Type Systems

Formally, Pure Type Systems are systems of rules about typing relations between terms. A typing relation: $A : T$ expresses that the term T is the type of the term A. The terms involved are like λ-terms in some generalised λ-calculus. These terms are constructed recursively using abstraction or application from variables and two or three constants, the *sorts* that are the only constants in a given PTS. In a typical PTS, we may have the sorts $*_s$ and $*_p$, and their type, the sort $\Box$. The sort $*_p$ corresponds to the type containing all possible propositions, and the type $*_s$ to the type containing all possible sets.

In order to calculate the type of a given term, the types of all free variables occuring in that term have to be given. Therefore, the type of an expression is always calculated in a certain *context*, which specifies the types of a number of variables. Formally, a context is a sequence: $x_1 : T_1, x_2 : T_2, \ldots, x_n : T_n$, where the x_n are variables and the T_n are terms that stand for the types of these variables. That a certain expression E has type T in a given context Γ is expressed by a *judgement*: $\Gamma \vdash E : T$. There are a number of inference rules that determine which judgements are valid. If a judgement is valid, the corresponding context has to be well-formed. Roughly, a context $\Gamma, x_n : T_n$ is well-formed, if Γ is well-formed (or empty) and x_n is a variable that is fresh in Γ, and T_n is a type which is *constructable* in Γ. This is the case, if $\Gamma \vdash T_n : s$ can be inferred, for some sort s [3].

A PTS can be used to model mathematical reasoning as a consequence of the socalled Curry-Howard isomorphism, the observation that propositions can be regarded as the types of their proofs: they will restrict the way in which proofs can be combined and ensure the soundness of the system. For instance, a proof f of $A \to B$ and a proof x of A can be combined (by application), to yield a proof $f(x)$ of B. Thus in PTSs proofs have explicit formal representations, and, because the judgements are decidable, they can even be checked mechanically. In short, PTSs allow us to represent both an agent's knowledge and his justifications for this knowledge. This is in accordance with the move towards 'intrinsic epistemics' as advocated in (Van Benthem 1991).

2.2 Interpretations

A context Γ can represent the 'body of hypotheses' of a mathematician. But to reflect the assumptions of an agent p about the real world, the context has to be grounded in his sense-impressions. This can be achieved if every inhabitable term T such that $\Gamma_p \vdash T : *_s$ or $\Gamma_p \vdash T : *_p$ has an *interpretation*. This interpretation is the personal ability of the agent p to judge whether something which is perceived is an inhabitant of T or not. Combining sense-impressions and interpretations, the agent takes certain types to

[3] In (Barendregt 1992) contexts are called 'pseudo-contexts' and well-formed contexts are called 'legal contexts'.

be inhabited. The justifications of the corresponding intuitive judgements do not admit analysis. They correspond to perceived objects, or direct physical evidence for a certain proposition.

2.3 Recognition of individuals

Like the ability to recognize types, the ability of an agent to recognize a particular object in the common environment is a subjective ability. In fact, the agent may not even recognize such an object consistently. In our model we assume that an agent exhibits only two 'degrees' of recognition:

> *Always recognize* an inhabitant when running into it in the environment. In that case the extension of the inhabitant is known by the agent (p) and we write $x\downarrow : T$ ($\in \Gamma_p$) in type theory.
>
> *Never recognize* an inhabitant when running into it in the environment. In that case the extension of the inhabitant is not known by the agent (p) and we write $x : T$ ($\in \Gamma_p$) as usual.

This is a somewhat crude distinction, but one that is defensible for the purposes of this paper. We call inhabitants of the first kind 'extensional objects' and those of the second kind 'intensional objects'.

2.4 Reasoning

Reasoning is one possible mechanism for knowledge growth. The reasoning of an agent is modelled by constructing new statements out of those occuring in a context, according to the inference rules. These rules characterize the notion of derivability '$\vdash$': $\Gamma \vdash E : T$ means that the inhabitant E can be constructed for type T on context Γ. Derivability on a given context reflects the agent's ability to find rational evidence (E) for an assertion (T) in his current knowledge state (Γ). We assume that the derivation rules are the same for all agents; knowledge states differ in content, but all agents 'use the same logic'.

After deriving a new conclusion, the knowledge state (Γ) cannot be updated directly by appending this statement ($E : T$) to it, since technically all statements in a context have to have a variable as term and E is a complex term. To allow the recording of conclusions, De Bruijn (1980) proposed to enrich the notion of context with 'definitions'. Using this idea, the context can be extended with a *definition* $x = E : T$ whenever a judgement $\Gamma \vdash E : T$ is derived in which E is a complex expression. This definition expresses that the variable x of type T may be used to refer to the complex term E in further derivations: $\Gamma, x = E : T \vdash x : T$. In other words, the definition 'abbreviates' the complex term with a fresh variable. At any point in time a definition can be 'unfolded' again, replacing the abbreviation (x) with the complex term E. In the presence of definitions, a well-formed context will look like this:
$x_1 : T_1, x_2 : T_2, x_3 = E_3 : T_3, x_4 : T_4, \ldots, x_n = E_n : T_n$. It represents a structured

collection of assumptions (atomic justifications), intermingled with conclusions (complex justifications) that have been drawn on the basis of these assumptions.

At first sight, recording the results of reasoning in the knowledge state may seem to be a mere ergonomical device: although the definition saves the trouble of going through the derivation of E again, the statement $E : T$ can be reconstructed on any context Γ' containing Γ ($\Gamma \subseteq \Gamma'$)[4]. Hence recording results of reasoning cannot help an agent to derive things that would have been 'unreachable' otherwise. But, in practice agents have limited deductive powers, allowing them only to oversee the more or less obvious consequences of their knowledge; conclusions which can be derived with a reasonable amount deductive work. To describe the deductive limitations of real-life agents, one can imagine a 'deductive horizon': an agent is able to immediately derive those consequences of his beliefs that lie within his horizon, all other consequence are 'out of sight'. For agents which have a deductive horizon, storing conclusions in the knowledge state literally broadens their horizon by bringing consequences into view that were unreachable before.

2.5 Knowledge from external sources

Whereas the reasoning process extends an agent's knowledge state from within, knowledge can also be obtained from external sources through communication and observation. Since all complex justifications reflect internal reasoning processes, the justifications of purely external information have to be atomic. Hence, this kind of information is represented by a context $y_1 : T_1, \ldots, y_m : T_m$, where $y_1, \ldots, y_m$ are fresh variables. In principle any context Δ could be appended to any well-formed context Γ to obtain a new context Γ, Δ. However, if this extension is to be meaningfull, the resulting context must again be well-formed. To this end we introduce the notion of an *extending segment* of a well-formed context:

Definition 1: A context Δ is an *extending segment* of a well-formed context Γ iff Γ, Δ is a well-formed context.

Given that the types represent concepts in the agent's knowledge state, this technical requirement captures the intuition that an agent can only extend his knowledge state with information expressed in terms of concepts that are already part of his knowledge state.

2.6 Common versus private knowledge

Central to a communication process between (two) agents is the continous extension of their common knowledge. To model the knowledge state of a dialogue partner, we therefore need to distinguish within the knowledge of each agent that part of its knowledge which it *assumes* to be shared. So for each agent p we have a context Γ_p which contains all its knowledge, within which we can distinguish a (sub)context Ψ_p, with $\Psi_p \subseteq \Gamma_p$, which contains all knowledge that, *according to p*, is shared.

[4] Formally, this property is known as the Start-lemma, see (Barendregt 1992

Both Γ_p and Ψ_p are well-formed type theoretical contexts in their own right. The common context Ψ_p is 'a part of' the private context Γ_p, and this relation can be defined in a straightforward way:

Definition 2: Given two legal contexts Γ and Γ', Γ is *a part of* Γ', notation $\Gamma \subseteq \Gamma'$, iff

1 for all statements of the form $x : T$ occurring in Γ either:
$x : T$ or a definition $x = E : T$ occurs in Γ',

2 all statements of the form $x = E : T$ occurring in Γ occur also in Γ'.

Under this 'part of'-relation, every definition in Γ must occur in Γ' (*2*), but declarations in Γ may be replaced by definitions in Γ' (*1*). This will be of use in Sect. 4, where it allows us to 'link' information in Ψ_p to information in Γ_p in a convenient way.

The distinction between common and private knowledge gives rise to more fine-grained accounts of reasoning and observation, which we discuss briefly in the following two paragraphs.

2.6.1 Reasoning

Since both Γ_p and Ψ_p are legal contexts, new statements can be constructed on either context using the derivation rules. Hence we can model the agent reaching 'private conclusions' in reasoning with private information (Γ_p) and 'common conclusions' in reasoning with common information (Ψ_p). Information that is shared with another agent is also privately available, this is reflected in the inclusion $\Psi_p \subseteq \Gamma_p$. It is not difficult to see that this inclusion guarantees that any statement derivable on an agent's common context (Ψ_p) is also derivable on his private context (Γ_p), but not the other way around[5]. According to Def. 2, Γ_p may not contain all statements of Ψ_p literally, since for some declarations $x : T$ in Ψ_p, Γ_p may contain a definition $x := E : T$. However, the rules governing the use of definitions in type theory ensure that $\Gamma_p \vdash x : T$ for each of these definitions. Hence, all statements occurring in Ψ_p are derivable on Γ_p, and so all derivations that are possible on Ψ_p are possible on Γ_p.

2.6.2 Observation of individuals

We can model the observation of an individual in the environment by an agent as an extension of the information state of this agent with an extensional object of the type of the observed entity ($x\!\downarrow : T$).

In most cases, an observation is 'private' in the sense that it need not be shared between the dialogue participants. The effect of such an observation is extension of the private context:

[5] It is possible to formalize the distinction between private and common information inside a single context, by regarding the two kinds of information as full-fledged epistemic modalities in the framework of Modal Pure Type Systems (see Borghuis 1994), a development of modal logic in typed λ-calculus.

Definition 3: In a *private observation* the observing agent (p), extends his personal context with a fresh extensional object of the type of the observed individual: $\Gamma_p \ \Rightarrow \Gamma_p, x\!\downarrow\!: T$.

Besides these private observations, there can also be 'shared' observations, necessary for resolving a reference of the dialogue partner by means of the common environment. Such references typically involve deixis: the agent points out a certain object in the domain to his counterpart. In these cases the result of the observation will be an extension of the common context of the observing agent (and that of the refering agent).

Definition 4: In a *common observation* the observing agent (p), extends his common context with a fresh extensional object of the type of the observed individual: $\Psi_p \ \Rightarrow \Psi_p, x\!\downarrow\!: T$, where the type T is given in some reference made by the dialogue partner.

3 Utterances

The previous section shows how the knowledge state of a communicating agent can be modelled by means of type theoretical contexts. In this section, we extend the model with a formal account of utterances: first the relation between the subjective knowledge states and the common language in which the agents communicate is discussed (Sect. 3.1), then we characterize the (communicable) content (Sect. 3.2) and pragmatic force (Sect. 3.3) of utterances.

3.1 Concepts and the common language

One the one hand, each agent has its own knowledge state, built on concepts which are meaningfull only to himself. On the other hand, the agents speak a common language in which they communicate. Hence each agent somehow connects its subjective concepts to words in this language. For instance, each of the agents will recognise a certain class of objects that are used by people to sit on. There is a word to describe this class in the language; in English objects in this class are called 'chairs'. There may be certain differences in interpretation, i.e. one agent may recognise an object as a chair that the other would call otherwise, but on the whole the two categories will match quite well. This is the case, because the use of all words is constantly being gauged by the language community.

We assume that for each agent (p) there exists a partial mapping $T_p \rightsquigarrow W$ from type variables in its knowledge state to words in the vocabulary of the shared language. How this mapping was formed (when the language was learned) is outside the scope of our model.

If a word in a given language corresponds to a concept, this concept is neccessarily common to a rather large group, i.e. the speakers of the language in question. The specific individual objects that we encounter in our daily life, such as 'my chair', are not commonly known among all speakers. Accordingly there exist no words that directly refer to these objects[6]. This means we have to refer to these objects as instances of a certain class, and try to point out the particular object through a description of characteristic properties that are accessible to the dialogue partner. If this is not possible, we have to content ourselves with introducing an arbitrary (indefinite) object of the given type in the dialogue (see Sect. 4.1).

The mapping between the knowledge states and the language in our model reflects this situation: it does not extend to the level of individuals and proofs (the inhabitants of inhabitants of $*_p$ or $*_s$). As a consequence, an agent cannot communicate every statement that is meaningfull within his knowledge state by purely linguistic means.

3.2 Messages

Against the background of our unsophisticated account of the relation between the knowledge states and the language spoken by the agents, we try to delineate the class of type theoretical expressions that correspond to dialogue utterances in the language. These expressions, which we call 'messages', are meaningfull in the knowledge states of *both* agents; they express content that an agent can communicate to his dialogue partner. To arrive at a definition of messages, we discuss a number of examples, based on the dialogue situation depicted in the table below.

Take a simple situation where both agents assume that they share all concepts related to their common vocabulary. This means that in the knowledge state of both agents, all types related to a word in the vocabulary (by $T_p \rightsquigarrow W$) are declared in their common context (Ψ_p). We assume a common vocabulary that consists of English words, hence our mapping constructs utterances in a sort of 'toy English', annotated sequences of words labelled with variables (like
'z : *lens*')[7]. In the table below, we see that agents A and B have a shared vocabulary that contains at least the words 'bundle','lens', 'primary' and 'enter'. The type variables corresponding to these words are declared in their common contexts Ψ_A and Ψ_B. Note that A and B have different type variables that correspond to the same word: e.g. in A's knowledge state the concept 'lens' is represented by the type x_3, in B's knowledge state it is represented by the type y_5. Depending on the situation, A and B may share more than just their vocabulary at the beginning of a dialogue. There may be certain general knowledge which they can correctly assume to share with their partner, or they may share certain information as a result of a previous conversation. This information will then also be represented in their common contexts. In the situation below, the following information is shared: apart from the types related to the vocabulary, each agent has a representation for a particular bundle (x_5 and y_{17} respectively) and a particular lens (x_6 and y_5 respectively):

[6] Exceptions exist: e.g. Earth.

[7] To construct a more realistic mapping, not only the vocabulary but the whole language should be taken into account. For a mapping from type theory to English, see (Mäenpää and Ranta 1990).

Ψ_A	$T_A \leadsto W$	Ψ_B	$T_B \leadsto W$
$x_1 : *_s,$	$x_1 \leadsto bundle$	$\ldots,$	
$x_2 : x_1 \to *_p,$	$x_2 \leadsto primary$	$y_5 : *_s,$	$y_5 \leadsto lens$
$x_3 : *_s,$	$x_3 \leadsto lens$	$y_6 : y_5,$	
$x_4 : x_1 \to x_3 \to *_p,$	$x_4 \leadsto enter$	$\ldots,$	
$x_5 : x_1,$		$y_{17} : *_s,$	$y_{17} \leadsto bundle$
$x_6 : x_3,$		$y_{18} : y_{17}$	
$\ldots$		$y_{19} : y_{17} \to y_5 \to *_p,$	$y_{19} \leadsto enter$
		$\ldots,$	
		$y_{32} : y_{17} \to *_p,$	$y_{32} \leadsto primary$
		$\ldots$	

Given this situation, we look at a number of cases where A wants to communicate something to B. We make a weak assumption about the things that A can say; we require that what he says is meaningfull on the common[8] part of his current knowledge state (Ψ_A). Technically this means that the content of what A says must be an extending segment of Ψ_A.

Consider the simple segment $z : x_1$. To A this means 'a bundle', it extends A's common context (z is Ψ_A-fresh). By the mapping $T_A \leadsto W$, it corresponds to the utterance 'z : *bundle*'. Upon hearing this utterance, B can decode it to the statement $z : y_{17}$ by means of his mapping $T_B \leadsto W$. Taking the variable z to be Ψ_B-fresh, this statement is an extending segment of B's common context as $\Psi_B \vdash y_{17} : *_s$ and $*_p$ is a sort. In other words, it is meaningfull to B.

In general, a statement or a segment that is meaningfull to A need not be meaningfull to B and vice versa, since there is no direct mapping from the inhabitants in their knowledge states to expressions in the language they share. For instance, the statement $z : x_4(x_5 x_6)$ which to A means something like 'bundle x_5 enters lens x_6' is not meaningfull to B. It corresponds to the utterance $z : \mathit{enter}(x_5 x_6)$. If A were able to utter this and B would try to decode it, B would end up with the statement $z : y_{19}(x_5 x_6)$. Taking the x's to be fresh, this is not an extending segment of B's common context ($\Psi_B \nvdash y_{19}(x_5 x_6) : *_p$). The statement is not meaningfull to B because it is unclear to him what the types of x_5 and x_6 could be. However, A could communicate something along the intended lines by the slightly longer segment: $u : x_1, v : x_3, z : x_4(uv)$ (with z, u, v Ψ_A-fresh). This segment meaning 'there is a bundle and there is a lens, and the bundle enters the lens' to A, corresponds to the utterance: $u : \mathit{bundle}, v : \mathit{lens}, z : \mathit{enter}(uv)$. If B 'decodes' this, he ends up with the segment: $u : y_{17}, v : y_5, z : y_{19}(uv)$ which *is* meaningfull to him[9]. Taking u, v and z to be fresh, the segment can be shown to extend B's common contexti because $\Psi_B \vdash y_{17} : *_p$, $\Psi_B, u : y_{17} \vdash y_5 : *_s$, and $\Psi_B, u : y_{17}, v : y_5 \vdash y_{19}(uv) : *_p$.

If we look at the differences between the short and the long segment, we see that the short segment has variables ocurring freely in the types that are not introduced in the

[8] Otherwise, it can never be meaningfull to B.

[9] Unfortunately A has not succeeded in communicating the identity of the objects involved. To solve this problem the apparatus introduced in Sect. 3 is required

segment itself and do not correspond to words in the vocabulary. The longer segment introduces all variables that do not correspond to words by itself, allowing B to give a meaning to all variables occurring in the types of the segment. Therefore, the idea presents itself that segments can only be meaningfull to both agents if they do not contain free variables in the sense described above. To make this idea more precise, we need the following definition.

Definition 5: A variable z *occurs free in a segment* Δ, $\Delta \equiv x_1 : T_1, \ldots, x_n : T_n$, iff z occurs free in T_i $(1 \leq i \leq n)$ and there is no statement $x_j : T_j$ with $1 \leq j < i$ such that $z \equiv x_j$.

If there are no free variables in an utterance, both agents can translate it to an extending segment Δ of their common context. So we can define messages, the class of segments that correspond to utterances in the following way:

Definition 6: Given the knowledge states, (Γ_p, Ψ_p), of the agents involved in a dialogue, each with a mapping $T_p \rightsquigarrow W$ of types in their knowledge states to words in their common vocabulary, an extending segment Δ of a common context Ψ_p is a *message* iff for all variables z occurring free in Δ there is a word W such that $z \rightsquigarrow W$.

3.3 Tags

Messages correspond to the content of utterances, and are the basic packets of information that the agents can exchange. Besides content, utterances in a dialogue have a function; the speaker intends the hearer to process the information expressed in the utterance in a certain way. In our model we take the pragmatic function to be a set of instructions to the hearer: (parts of) messages are tagged with 'pragmatic clues', which tell the hearer how they should be interpreted on the knowledge state and how he should react.

The most important of these clues is the 'polarity' of the message, which tells the hearer whether the extending segment is to be processed 'passively' or 'actively'. So far we have described only one way of dealing with extending segments (Sect. 2.5): the agent simply appends the extending segment to one of the two contexts constituting his knowledge state. This is passive in the sense that the agent makes no effort whatsoever to connect the new information represented by the extending segment to the information already present in his knowledge state. The new information is simply stored as a set of additional 'hypotheses' or 'assumptions' (all justifications in the segment are atomic). However, the agent can digest new information in a more active way by trying to find justifications (objects and proofs) in his own current contexts to replace the dummy inhabitants of the statements in the extending segment. If he succeeds in doing so, we say that the agent has constructed a 'realization' for the new information in his original context.

Definition 7: let $\Delta \equiv x_1 : T_1, ..., x_n : T_n$ be an extending segment of Γ, and let
$\Gamma \vdash D_1 : T_1$ and
$\Gamma \vdash D_2 : T_2[x_1 := D_1]$, and
$\Gamma \vdash D_3 : T_3[x_1 := D_1, x_2 := D_2]$, and

... and
$\Gamma \vdash D_n : T_n[x_1 := D_1, ..., x_{n-1} := D_{n-1}]$ then we call
$\Delta^* \equiv x_1 = D_1 : T_1, ..., x_n = D_n : T_n$ a *realization* of Δ in Γ under the substitution $[x_1 := D_1, ..., x_n := D_n]$.

The active agent then appends the realization Δ^* to his context instead of the extending segment Δ. This does not provide the agent with new information, but the point is that active segments act as selective 'hooks' with which a message can be connected to existing knowledge on the part of the hearer. Formally, the fact that active processing of segments does not yield new information is reflected by the following theorem[10], which shows that realizations can be eliminated:

Theorem 1: Assume $\Gamma, \Delta \vdash B : C$ Let Δ^* be a realization of Δ in Γ under the substitution $[S]$, then $\Gamma, \Delta^* \vdash B : C$ and $\Gamma \vdash B[S] : C[S]$.

Hence an extending segment Δ of a context Γ can be used in two quite different ways: either as *hypothesis* extending the current knowledge state, or as a *requirement* for which a realization is to be constructed in the current knowledge state. We call the former use of segments 'positive', the latter 'negative'. As we will see, this polarity determines the direction of the flow of information in our communication model. Formally we express the polarity of messages in the following way:

Definition 8: A (part) of a message $\Delta \equiv x_i : T_i, ..., x_j : T_j$ is *positive*, when it is to be appended to one of the contexts in the knowledge state of the agent receiving it. Notation: $x_i : T_i, ..., x_j : T_j \ [x_i, \ldots, x_j]^+$.

A (part) of a message $\Delta \equiv x_i : T_i, ..., x_j : T_j$ is *negative*, when a realization Δ^* of it is to be constructed on one of the contexts in the knowledge state of the agent receiving it. Notation: $x_i : T_i, ..., x_j : T_j \ [x_i, \ldots, x_j]^-$.

Note that a message can be tagged in such a way that different parts of it have different polarities, instructing the receiving agent to add the statements in the positively tagged parts and to construct realizations for the statements in the negatively tagged parts.

Given the polarity of (parts of) messages, we will distinguish two more tags that specify further instructions to the hearer. The first one indicates the part of the knowledge state to which the segment is to be added, or on which a realization has to be constructed; the private or the common context. The second one indicates which part (if any) of the realization of a segment constructed by the hearer is to be returned in response to the speaker's utterance.

[10] This theorem is simply an iterated version of the 'Substitution Lemma' for PTSs, see (Barendregt 1992).

4 Communication

In this section we show how the machinery introduced in the previous sections can be used to perform fundamental communicative transactions, like referring to objects, resolving object references, presenting information to, and obtaining information from a dialogue partner.

4.1 Describing objects

In order to designate a certain concrete object to another agent (B), agent A must give an individuating description of the object, using only resources that are common between him and B. These resources are

ψ_A: the common context of A, which contains shared general knowledge as well as information that was already exchanged in the present dialogue,

χ: the common environment of A and B, in which the agents can use deixis to refer to objects.

A's individuating description takes the form of a negatively tagged message $\Delta\ [\ ]^-$, for which B must find a realization using the resources he has in common with A: Ψ_B and χ. This message Δ consists of a declaration $z : T$, where z is a (fresh) variable of the type (T) of the object that is to be designated, possibly followed by other statements expressing further distinguishing properties of this object. If B succeeds in finding a realization, he extends his common context (Ψ_B) with it. Agent A, who already had a realization for the message, updates his common context (Ψ_A) in a similar way. Hence the object designated by A becomes an object that is 'shared' between A and B.

If it is not possible for A to give an individuating description using the common resources, he must contend himself with introducing a new 'anonymous' object of the intended type (T). This can be done by means of a positively tagged message $z : T[z]^+$, which will result in an update of the common contexts of both A and B with this new object.

4.2 Presenting information

The polarity of messages that is vital to the description of objects is equally important on the level of complete dialogue utterances, because it specifies whether an utterance is intended to present information ('assertion') or to obtain information ('question'). As we will see, 'questions' provide the hearer with a requirement for which he must construct a realization on his knowledge state (an 'answer'). 'Assertions' provide the hearer with information extending his knowledge state. The pragmatic clues instructing the hearer how to process such a complete utterance can be more complex and more detailed than those for the description of objects; the corresponding messages have a richer repertoire of tags.

The most important refinement in the instructions to the hearer is a tag indicating on which part of his knowledge state a (part of a) message of a given polarity is to processed. With descriptions all messages were processed on the common context of the hearer, on the level of utterances the private context also comes into play. This is essential, since one of the main epistemic effects of communication is that information that is initially privately available to one of the participants becomes shared between all participants. Hence (parts of) messages will not only be tagged as positive or negative, they are also marked with the part of the knowledge state to which they are to be added or on which a realization for them is to be constructed. This is indicated by subscripts $[\ ]_\Gamma$, for the private context, and $[\ ]_\Psi$, for the common context.

A speaker uses an assertion-type utterance to provide the hearer with information that a certain state of affairs is the case. If the utterance is felicitous, this information is new to the hearer and will be accepted by him. To introduce this new 'evidence' into the knowledge state of the hearer, the relevant (part of the) message has to be tagged positively. Since the assertion was uttered publicly, a further pragmatic clue instructs the hearer to extend his *common* context (Ψ_B) with this positive message. Likewise, the speaker should record that the information that was (in part) private has now become public by updating his common context (Ψ_A). Hence the net effect of an assertion on the knowledge states of hearer and speaker is that the common contexts of both are extended with a statement signifying that the proposition expressed by the assertion is inhabited.

Although the update with a positive message containing an inhabited proposition is characteristic of utterances which present information, the overall structure of a message corresponding to such an utternace can be more complex because the proposition may refer to objects in the domain in any of the ways discussed in the previous section. For example, consider the assertion 'The primary bundle enters a lens'. It might correspond to the tagged message[11]:

$$(1) \qquad b : \mathit{bundle}, p : \mathit{primary}(b), l : \mathit{lens}, q : \mathit{enter}(b, l)\ [b, p]^-_\Psi\ [l, q]^+_\Psi$$

The tags show that this message (Δ) has a positive and a negative part. The first, negative, part (Δ_1) corresponds to the definite noun phrase 'the primary bundle':

$$(2) \qquad b : \mathit{bundle}, p : \mathit{primary}(b)\ [b, p]^-_\Psi$$

It instructs the hearer to find a realization for b and p on his common context Ψ_B. Assuming that the speaker's use of a definite description was appropriate, B will find an object, say y_{34}, representing the bundle in his common context along with a proof that it is the primary bundle, say N. The hearer extends his common context with this realization,

[11] In this example as well as the ones that follow, we abstract form the differences in type variables between the knowledge states of A and B, by using words from the common vocabulary W as types in messages. In doing so, we deliberately blur the distinction between message and utterances introduced in Sect. 3.2, because we want to concentrate on the level of inhabitants in these examples. Also we write '$\mathit{enter}(b, l)$' instead of the Curryed '$((\mathit{enter}\ b)l)$', the notation commonly used in λ-calculus.

$b := y_{34} : \mathit{bundle}, p := N : \mathit{primary}(b)$, and proceeds by processing the second part of the message. This part (Δ_2) contains the proposition asserted by the speaker that for some lens l the primary bundle enters l:

(3) $$l : \mathit{lens}, q : \mathit{enter}(b, l)\ [l, q]^{+}_{\Psi}$$

Note that the b that occurs free in Δ_2 is now bound in the extended common context by the definition $b := y_{34} : \mathit{bundle}$. The second part of the message is tagged positively; the hearer is supposed to absorb the information 'passively' by adding the statements in Δ_2 to his common context. The first statement in Δ_2 is an indefinite description which introduces a new lens (l) into the common context. The second statement introduces a new 'piece of evidence' into the common context, a proof object (q) for the proposition $\mathit{enter}(b, l)$.

The processing of the entire utterance therefore updates the common context of the hearer in two steps: $\Psi_B \Rightarrow \Psi'_B$ with $\Psi'_B \equiv \Psi_B, \Delta_1^*$ where Δ_1^* is a realization of Δ_1 in Ψ_B under substitution $[S]$, followed by $\Psi'_B \Rightarrow \Psi''_B$ with $\Psi''_B \equiv \Psi'_B, \Delta_2$. According to theorem 1, Δ_1^* can be eliminated in favour of $[S]$ yielding $\Psi''_B \equiv \Psi, \Delta_2[S]$ (where $\Delta_2[S]$ abbreviates the application of $[S]$ to the statements in Δ_2). From this point of view the net effect of the entire message on the common context of the hearer is an update with evidence for the proposition that for some lens (l) the primary bundle (y_{34}) enters that lens.

It should be noted that the reaction of the hearer in this example is the simplest or 'most cooperative' one possible; he adds the information provided by the speaker without questioning it in any way. Depending on factors in the dialogue situation not considered here, this reaction could be more 'cautious'[12].

The succesfull utterance of the assertion by A not only affects the knowledge state of B, but also that of A himself. The utterance was made in public, and hence affects the common context of A in the same way as the common context of B. As we described above for B, this results in an extension of the common context: $\Psi_A \Rightarrow \Psi''_A$ with $\Psi''_A \equiv \Psi_A, \Delta_1^*, \Delta_2$, or equivalently $\Psi''_A \equiv \Psi_A, \Delta_2[S]$ (where $[S]$ substitutes A's representation of the primary bundle for b).

Privately the speaker will know more, since he must have some justification for his utterance. In particular, he must have 'evidence' for the proposition expressed asserted: the positively tagged part of the corresponding message, Δ_2. The least we can assume about this evidence is that there exists a realization of Δ_2 on his private context Γ_A, e.g.: $\Delta_2^* \equiv l := x_{35}\!\downarrow : \mathit{lens}, q := M : \mathit{enter}(b, l)$. This realization shows in which respects the speaker can know more in his private context than in his common context: in Γ_A he knows which lens the bundle enters ($x_{35}\!\downarrow$), rather than just 'a lens' (l) in Ψ_A. Moreover, in Γ_A he has a structured proof (M) for this rather than a 'dummy' (q) in Ψ_A. Agent A can connect his 'private' justifications to his 'common' justifications by updating his private context with the realization: $\Gamma_A \Rightarrow \Gamma'_A$ with $\Gamma'_A \equiv \Gamma_A, \Delta_2^*$. By adding these definitions,

[12] For instance, the DenK-system will not accept all assertions made by the user, because it is an expert on the domain whereas the user is a novice.

$x_{35}{\downarrow}$ is linked to l and M to q (b was already linked to A's representation of the primary bundle by the update of Ψ_A with Δ_1^*). Without this link, A would be unable to combine information about l and $x_{35}{\downarrow}$ in his private context.

4.3 Obtaining information

In the type theoretical perspective, specifying to an agent what you want to know corresponds to presenting him with a negative message; an extending segment for which he has to construct a realization on his knowledge state. If the speaker's request is to be felicitous, he himself does not have the requested information, i.e. there is no realization for the negative message on his common or private context. In view of the similarity between the common contexts of the dialogue participants, there is no realization on the common context of the hearer. Therefore the hearer must construct the realization on his *private* context, allowing him to use all information available in his knowledge state. So in our model, questions correspond to annotated messages which characteristically have a negative part that is to be processed on the private context ($\Delta\ [\ ]_\Gamma^-$).

Once the hearer has constructed a realization, he has to return the information in some answer or other. The speaker can influence this reaction by the kind of question he asks. For instance, if he asks a 'Yes/No-question', he is only interested in whether or not the hearer has succeeded in constructing a realization, not in its content. On the other hand, if he asks a so-called 'Wh-question' (What,Who,Where, ...), he wants to know the content of a specific part of the realization constructed by the hearer. For a type theoretical treatment of the various question-answer relations the reader is referred to [Piwek 1996]. In this section, we discuss the basic epistemic effects of questions by means of Wh-questions.

In Wh-questions, the Wh-constituent (e.g. 'Which lens ...?') specifies what information that the speaker expects the hearer to provide in the answer. In our type theoretical perspective this means that in the message corresponding to the question, one variable is pragmatically marked; the hearer is supposed to return the realization he constructs for the variable on his private context. This marking takes the form of an additional tag: in the annotation of the negative (part of the) message that is characteristic for the question, the variable corresponding to the Wh-constituent is underlined ($\Delta\ [\ \ldots, \underline{x}, \ldots\ ]_\Gamma^-$).

We discuss the epistemic effect of a Wh-question by means of an example related to the one we gave earlier for assertions (Sect. 4.2). Suppose the speaker asks the question 'Which lens does the primary bundle enter ?'. The message (Δ) corresponding to this question is the same as that for the assertion 'The primary bundle enters a lens'. However, the pragmatic clues which instruct the hearer how to process the message are different:

(4) $$b : bundle, p : primary - b, l : lens, q : enter(b, l)\ [b, p]_\Psi^-\ [\underline{l}, q]_\Gamma^-$$

The first part of the message (Δ_1) is tagged as before:

(5) $$b : bundle, p : primary(b)\ [b, p]_\Psi^-$$

Hence it will be processed by the hearer in the way described above, yielding an update of Ψ_B with Δ_1^*; a realization for 'the primary bundle'. The tags for the second part of the message (Δ_2) differ from those in the assertion-example in a number of ways:

$$l : lens, q : enter(b, l)\ [\underline{l}, q]^{-}_{\Gamma} \tag{6}$$

Firstly, the Δ_2 now has a negative polarity, instructing the hearer to view it as requirement for information rather than a set of working hypotheses. Secondly, Δ_2 has to be processed on the *private* context. In other words, the hearer is required to construct a realization for Δ_2 on Γ_B; he has to find a lens and construct a proof that the primary bundle enters this lens[13]. We assume that the hearer is able to construct such a realization, say $l := y_{79}, q := M : enter(l, b)$. At least one of the items in this realization must be 'strictly private' in the sense that it is available on Γ_B but not on the subcontext Ψ_B, for the reason discussed above: if the entire realization could be constructed on Ψ_B, a realization could also be constructed on Ψ_A and then A's question would not be felicitous. In this particular case, the lens (y_{79}) could be in Ψ_B but the proof object cannot be derived on Ψ_B. The update of his private context with the realization, $\Gamma_B \Rightarrow \Gamma'_B$ with $\Gamma'_B \equiv \Gamma_B, \Delta_2^*$, brings B in a position where he (privately) posseses all information needed to answer A's question.

The underlining of l, tells B which part of the realization A is interested in. This tag indicates that in answer to the question B has to return his realization for l to A; A wants to know the identity of the lens that is entered by the primary bundle. Since the identity of objects cannot be communicated directly, B will have to describe the lens y_{79} to A using common resources in one of the ways explained in 4.1. For instance, if it is commonly known among A and B that the microscope contain a number of condensor lenses which are arranged in some order, B could use the assertion 'The primary bundle enters the first condensor lens' to describe the lens to A. This assertion will update the common contexts of both A and B with the answer to A's question (cf. Sect. 4.2).

Uttering an information request has no direct effect on the knowledge state of the speaker: since we model updates as an extension of a knowledge state with positive or realized segments, the knowledge state of the speaker does not change until the hearer answers the question and the speaker can update his common context (Ψ_p) with the answer. Hence we cannot model the *posing* of a question at this level. In the 'Which-example' above, this is no problem, since the hearer gives a full answer: 'The primary bundle enters the first condensor lens'. Processing this answer will update the knowledge state of A correctly without the need to go back to the question in any way. In general this is not the case; answers can be incomplete ('The first condensor lens'), requiring the speaker to 'keep the question in mind' in order to obtain the correct update[14].

[13] Type theoretically this requirement is well-formed, Δ_2 is an extending segment of Γ_B: the variable b occurring free in Δ_2 is bound in Γ_B after the update of Ψ_B with Δ_1^* because of the inclusion $\Psi_B \subseteq \Gamma_B$.

[14] In the DenK-system, such 'memory' is indeed provided. In fact it has a more general use, since a similar problem arises when an utterance has been partially interpreted, but further clarification is needed to deal with an unresolved referring expression in this utterance.

5 Conclusions

We have presented a simple model of communication processes, where we limit ourselves to the case where two participants exchange information about a common environment. The model is based on an explicit formalization of the the knowledge states of the dialogue participants in PTSs, which stresses the private and subjective nature of these states, and provides a formal criterion to decide which extensions of these subjective knowledge states are meaningfull. Modelling a common vocabulary as a mapping from types to words, one can define the set of messages: the subset of meaningfull extensions of a knowledge state that can be translated with the given vocabulary. These messages can not only be used to add information to a knowledge state, but also to extract information from it, due to the notion of 'realization' (see Def. 7).

A few pragmatic tags which, given this model, are quite obvious, allow fundamental communicative transactions to be realized through tagged messages. Given the simplicity of the model, it is remarkable how closely these tags, which were operationally motivated, mirror important distinctions that figure in descriptive theories, like definite vs. indefinite, and assertive vs. interrogative.

We feel that the model could provide a point of departure for a computational theory of dialogue. We are strengthened in this conviction by the mutually compatible explanations of various linguistic phenomena that have already been formulated in this framework, such as: presuppositions (Krahmer and Piwek 1997), the resolution of definite descriptions (including anaphora and uses of deixis, (Beun and Kievit 1996)) and question/answer relations (Piwek 1997).

References

Ahn, R.: 1994, Communicating contexts: A pragmatic approach to information exchange, in P. Dybjer, B. Nordström, and J. Smith (eds.), *The proceedings of the BRA workshop: Types of Proofs and Programs*, Vol. 996 of *Springer Lecture Notes in Computer Science*, Springer Verlag, Berlin

Ahn, R. et al.: 1994, The denk-architecture: A fundamental approach to user-interfaces, *Artificial Intelligence Review* 5(8)

Barendregt, H.: 1992, Lambda calculi with types, in S. Abramsky, D. Gabbay, and T. Maibaum (eds.), *Handbook of Logic in Computer Science*, Oxford University Press, Oxford

Beun, R. and Kievit, L.: 1996, Resolving definite expressions in denk, in *Proceedings of the 5th International Pragmatics Conference*, Mexico City

Borghuis, T.: 1994, *Coming to Terms with Modal Logic: on the Interpretation of Modalities in Typed λ-calculus*, Ph.D. thesis, Eindhoven University of Technology, Eindhoven

De Bruijn, N.: 1980, A survey of the project automath, in J. Seldin and J. Hindley (eds.), *To H.B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalisms*, pp 589–606, Academic Press

Krahmer, E. and Piwek, P.: 1997, Presupposition projection as proof construction, in H. Bunt, L. Kievit, R. Muskens, and M. Verlinden (eds.), *Proceedings of the Second International Workshop on Computational Semantics*, pp 122–134, Tilburg

Mäenpää, P. and Ranta, A.: 1990, An implementation of intuitionistic categorial grammar, in L. Kálmán and L. Pólos (eds.), *Papers from the Second Symposium on Logic and Language*, Akademiai Kiado, Budapest

Piwek, P.: 1997, *The Construction of Answers: A Deductive Approach to Answerhood*, This volume

Van Benthem, J.: 1991, Reflections on epistemic logic, *Logique et Analyse* 133-134

On First Order Information Exchange*

Paul Dekker

ILLC/University of Amsterdam

1 Introduction

Our main interest in this paper lies with the felicitous exchange of first order information. We are concerned with the question what licenses a speaker to convey open propositions or to utter sentences with pronouns. Although Grice's maxims of quality seem to give an appropriate answer to the question what licenses the felicitous exchange of propositional information, and although first order notions of truth and support have been developed within theories of discourse representation and dynamic interpretation, the question about first order licensing has to our opinion remained unanswered sofar.

In the course of this paper we argue for a specific formal notion of licensing which satisfies some intuitively plausible principles concerning first order information exchange. According to the view exposed in this paper, speakers may convey information about subjects which have been introduced in a discourse only if these correspond to subjects they themselves have information about. It makes no substantial difference if these subjects have been introduced in the discourse by the speakers themselves or by other agents. We will argue that indeed the last type of phenomenon, which has become known as "cross-speaker anaphora", is not structurally different from the phenomenon of ordinary, single speaker, anaphora. The main and rather trivial difference between single and cross-speaker anaphora resides in the fact that one agent may simply fail to have information about (the identity of) subjects introduced by other agents. Failing such information, it would clearly be inappropriate for him to exchange information about them.

We will proceed as follows. In section 2 we will critically discuss two preliminary sketches of the phenomenon of cross-speaker anaphora, one by Francez and Berg, and one by Groenendijk, Stokhof and Veltman, and we will argue that these analyses do not make sufficiently clear when or why cross-speaker anaphora is licensed. Then, in section 3, we will argue that, despite appearances, it is not even clear yet what licenses single speaker anaphora, and we develop a general notion of support which can be conceived of as a first order statement of Grice's maxims of quality. Then, in section 4, we will show that the principles governing single speaker anaphora also can be used to describe the licensing of

* This note reports about work in progress on models of information exchange. I wish to thank Nicholas Asher, Jelle Gerbrandy, Jeroen Groenendijk, and Ede Zimmermann for helpful comments. The research for this work is supported by a fellowship from the Royal Netherlands Academy of Arts and Sciences (KNAW), which is gratefully acknowledged. Needless to say that all errors are mine.

cross-speaker anaphora. Section 5 concludes with some considerations of a more general nature.

2 Cross-Speaker Anaphora

In this section we will discuss some observations about cross-speaker anaphora which have been made in (Francez and Berg 1994) and (Groenendijk et al. 1997).

Multi-Agent DRT

For as far as we know, Nissim Francez and Jonathan Berg have been the first to pay substantial attention to the phenomenon appropriately dubbed "cross-speaker anaphora" (Francez and Berg 1994). Cross-speaker anaphora is a phenomenon occurring in multi-agent dialogue, "whereby one speaker refers anaphorically to a discourse entity introduced by *another* speaker." Here is an example:

(1) A: A man is walking in the park.
(2) B: He whistles.

Cross-speaker anaphora resembles so-called "donkey-anaphora", with the striking difference that the person who uses the anaphoric pronoun is different from the one who utters the antecedent (indefinite) noun phrase. This phenomenon raises a couple of questions, one which Francez and Berg call the 'commitment problem'. In their own words: "to what conditions about a discourse entity is a speaker committed when using cross-speaker anaphora?"

Francez and Berg present a first rough analysis of cross-speaker anaphora which has three distinctive aspects. First, for a proper representation of multi-agent dialogue, the various commitments of different agents are kept separate. Employing the representational format of discourse representation theory (*DRT*, henceforth), the various contributions of different agents are not lumped together in one cumulative representation, but each agent is assigned a separate section in the *multi-agent discourse representation* where his contributions are evaluated. The second feature of Francez and Berg's analysis is concerned with the representation of cross-speaker anaphora. If one agent B uses a pronoun y to refer back to a discourse marker x introduced by another agent A, then B's section in the multi-agent discourse representation is updated with the condition that y depends on x. The third aspect of Francez and Berg's analysis relates to the evaluation of such a dependency condition. If someone refers to a discourse entity introduced by another agent, then he is committed to (the satisfaction of) conditions "directly arising" from the introduction of the antecedent. This is called the "core condition commitment". The core condition includes the core description under which the relevant discourse entity is introduced

Our interest in this paper lies in the semantic (and pragmatic) impact of cross-speaker anaphora, and we think Francez and Berg's core condition commitment (*CCC*, henceforth) is both too rigid and too weak in this respect. IThe *CCC* inflexibly associates a descriptive content with any discourse entity, which any embedding of the entity must respect. This is too strong a requirement, because one can deny a discourse entity to have any property which other agents have ascribed it, this without contradicting oneself. Consider:

(3) *A*: Yesterday, Tom went to the Zoo with the youngest son of a colleague of his. The boy enjoyed it very much to spend the day with him.
(4) *B*: Sure, because '*he*' as a matter of fact was a nineteen year old *girl*, and because they went dancing, instead.

In the first example A introduces an entity in the discourse as a young son of a certain colleague of Tom. B apparently objects to this description, stating that it was in fact a nineteen year old girl whom Tom went out with. We think B may successfully refer to the entity introduced by A, but, clearly, he, thus, has not committed himself to the condition that the entity is both a young son of a colleague of A and a nineteen year old girl.

One might, of course, object that in the present example the core condition is not "x is the youngest son of a colleague of Tom", but "Tom went to the Zoo with x", which B is committed too. This might serve as an alternative explanation of the above example, but it won't do in general. Consider N's reply (6) to R's utterance of (5):

(5) *R*: A man is sleeping on a park bench over there.
(6) *N*: It is not a man, it is a woman, and she is not asleep, she is just sunbathing. Besides, it is not a parkbench.

N here denies almost *all* conditions on the discourse entity introduced by R. If we strip of the attributes which N denies it to have, the entity introduced by R is seen to be no more than just an x such that x is on a y over there. It seems this can hardly count as *the* core condition by means of which sentence (5) introduces its subject. Again, one might argue that it is not sentences that inflexibly associate their subjects with core conditions, but that it is utterances that do, but then it seems we are left without any analysis. Notice, for instance, that N might have *thought* "besides, it is not a parkbench", and *not* have *said* so. Thus N does not need to be committed to conditions on a discourse entity imposed by other agents, also if he doesn't explicitly deny them.

Our main objection to the *CCC*, however, is that it does not at all do justice to the intuition that cross-speaker anaphora involve some notion like talking about the same thing. In the little dialogues above A and B, and R and N seem to be speaking about the same thing, and it even seems that the corrections made by B and N only make sense if they are assumed to be speaking of the same entity as A and R are. However, nothing in the *CCC* requires that the two are speaking of the same individual. According to the *CCC*, reference to the same discourse entity by different agents does not at all constrain these to invoke the same individual, it at most requires the referents to be a member of the set of entities satisfying the core condition. We think Francez and Berg thus fail to account for an important characteristic of cross-speaker anaphora.

Multi-Speaker Update Semantics

Jeroen Groenendijk, Martin Stokhof and Frank Veltman have presented a further analysis of the phenomenon of cross-speaker anaphora, and their main conclusion is that it is not at all such an unconstrained phenomenon as Francez and Berg's analysis seems to suggest (Groenendijk et al. 1997). *GSV* discuss examples like the following:

(1) A man is walking in the park.
(7) He is wearing blue suede shoes.

Everybody will agree that there is nothing strange with an utterance of (7) after an utterance of (1) by one single speaker (or in a multi speaker *monologue*). However, *GSV* observe, there seems to be something marked if, without further context, one agent, A, utters (1) and another, B, replies with (7). If it is not clear whom exactly A is talking about, or which story he is about to tell, then B's reaction indeed is pretty odd.

This is not to say that B is never allowed to refer back to A's subject. For instance, if (1) is understood as a report about a directly visible situation, then A may be referring to a particular individual. By replying with (7) B then may observe that this man is wearing blue suede shoes. And also if B doesn't know whom A is talking about, or which story he is about to tell, B may ask questions about this subject, and draw conclusions about him on the basis of what was said. So, after A has uttered (1), a reaction with any of the following two sentences would be perfectly alright:

(8) Is he wearing blue suede shoes?
(9) He must be wearing blue suede shoes then.

Cross-speaker anaphora appears to be possible if the relevant antecedents are 'definite' enough or when the pronouns are used with non-factual or non-assertoric force (and also in the case of multi-speaker monologue, when different speakers are said to act as a single agent).

GSV go some way in analyzing the second kind of cross-speaker anaphora in terms of a distinction between "direct" and "indirect" information. If people have direct information then it may be expressed straightforwardly, but if it is indirect information then it should be expressed using appropriate (modal) qualifications. Thus, when B doesn't know which individual A's utterance of (1) is about, then he at best has indirect information that that individual is wearing blue suede shoes. In that case he would not be licensed to utter (7) and he would only be licensed to utter (9) then. There is a certain appeal to this analysis, although it raises a couple of questions.

In the first place, it is not clear what is exactly meant with the terms "direct" and "indirect". The use of these terms invokes the suggestion that what is at stake here is a distinction between information for which we have direct, perceptual, evidence, as opposed to information obtained by other means, from hearsay, readwrite, watchtelevision, inference and so on. But this cannot be the relevant distinction it seems. Consider again B's reaction to A's remark that Tom went to the Zoo with a colleague's son. Clearly B's source may be other than perceptual, for instance plain gossip, tabloids, or just good thinking. Still there seems to be nothing basically wrong with his referring to A's subject without any modal qualification.

We think, that what *is* crucial is that one's utterances are supported by one's own information, not that of the other participants. If one's utterance, like B's (9) above, depends on information provided by other participants, or even refers to subjects introduced by other participants, then one has to add a modal qualification to indicate that one does not have independent support oneself. Thus, B's reply (9) should be understood as referring to the information expressed by A's utterance, and as saying that if indeed a man is walking in the park, then he is wearing blue suede shoes. In other words, anyone who buys A's claim that a man is walking in the park—B says—may buy from him that he wears blue suede shoes.

In the second place, the distinction between information that has to be modally qualified, and information that need not, at least involves a distinction between the information which one agent has from the information expressed by others in a dialogue (cf., Groenendijk et al. 1997, p. 24). But this seems to force us to revaluate the update semantic notion of meaning, for this distinction can only be maintained, it seems, if people do *not* in general update their information with what other agents say. An alternative dynamic conception of meaning which may be appropriate in this respect is one presented in (Beaver 1995), where the interpretation of a sentence is spelled out in terms of updates of what may be the speaker's picture of the common ground. In this paper we want to adopt a less involved view. We will stick to the fairly standard practice of assigning contents to information states and sentences alike. In our overall view, the dynamic resides in the combinatorics of these 'information containers'. Thus it is a straightforward matter to keep (the contents of) the sentences uttered by one agent, as well as his own information, separate from those of others.

The Definiteness Constraint

GSV also consider cross-speaker anaphora involving definite antecedents. When protagonist A introduce a definite subject, respondent B may be able to identify him. Because the respondent may have and convey information he himself has concerning the individual which (he thinks) his protagonist has introduced, the use of anaphoric pronouns in such cases can be seen to be licensed without requiring modal qualification. GSV also think of indefinite noun phrases as sometimes supplying definite antecedents. That is, they propose an analysis of sentence (1) as (10):

(1) A man is walking in the park.
(10) T is a man who is walking in the park.

where 'T' is a demonstratively referring phrase or a definite denoting phrase, which then may serve to license a pronoun by other agents.

Again we think there is both something promising and something problematic about GSV's proposal. If an indefinite licenses non-modally qualified cross-speaker anaphora, this may have to do with a definite or specific interpretation of the indefinite, but a specific analysis of indefinites can be built upon their ordinary dynamic semantic analysis by pragmatic means.

We also think it would be inappropriate to require a re-analysis of indefinite noun phrases as definite noun phrases, in order for them to license cross-speaker anaphora. Consider the following dialogue:

(11) H: A magistrate from Gotham village has confessed battering young girls.
N: They say he suspected them of sorcery. Do you know if more magistrates confessed?
H: I don't know.
N: Do you know who he is?
H: No idea, he preferred to remain anonymous.

In a situation liuke this, H and N would probably agree that they talk about one and the same subject, perhaps one they read about in the (news-)paper. And probably they can come up with some identifying description when asked *whom* they talk about, e.g., "the magistrate reported about in *The Daily Planet* two days ago." So, the above dialogue can be seen to involve an exchange of information about one and the same subject. However it is not at all clear from the dialogue which individual is concerned, or which definite must be used to refer to it. (There is a sense in which even H and N can be said not to know which individual it is.) But clearly we can understand H and N's utterances, without knowing which definite term, if any, is to replace the indefinite term "a magistrate of Gotham village". In other words, we can make perfectly good sense of the dialogue, without knowing how the example should be (re-)analyzed on *GSV*'s account.

To conclude, we think with *GSV* that non-modally qualified statements, made with or without pronouns, should provide information which the speakers themselves dispose of, and we also think that this invokes a requirement that they report on definite subjects they themselves have information about. We do not, however, think this excludes subjects introduced in the discourse by other agents by means of indefinite noun phrases.

3 Modeling Aboutness

In the next section we will argue that the principles governing the licensing of cross-speaker anaphora follow from principles governing the licensing of utterances more in general. But before we can give substance to that claim, we have to say what these principles are, and then it appears that the principles governing the licensing of *single* speaker anaphora are not yet sufficiently understood.

Licensing Open Propositions

When talking about the licensing of utterances, or about the justification for exchanging certain information, one tends to think of Grice's maxims of quality: do not say what you believe to be false and do not say that for which you lack adequate evidence. As long as we are dealing with ordinary propositions in a possible worlds framework, it seems easy enough to spell out what this kind of licensing amounts to. If s is the set of worlds compatible with one's information, then s should be a subset of p in order for one to be licensed to state that p. However, it appears to be not too straightforward to apply this notion of licensing in a first order framework. What should we think of as licensing the utterance of open formulas, formulas containing one or more free variables, or, if we think more in terms of natural language, sentences containing pronouns?

Two options for ruling the utterance of an open sentence licensed immediately spring to mind, and we can reject them just as straightforwardly. They involve the universal and the existential closure of the proposition. According to the first, one is licensed to state that Wx, or to utter "He whistles" only if one has the information that $\forall x\, Wx$, or that everything whistles. According to the second, one is licensed only if one has the information that $\exists x\, Wx$, or that something whistles. Clearly, this doesn't make much sense. The universal requirement is far too strong, the existential far too weak.

A mediating alternative consists in requiring that one has the information that the formula 'Wx', or the sentence "He whistles", holds of the individual referred to by means of the variable 'x', cq. the pronoun "he". This clearly makes much more sense than the two previous options, but it still is not appropriate. As we have seen above, in the dialogue about the Gotham village magistrate, it need not be clear which individual an agent refers to with a pronoun, not even for him- or herself. Clearly this does not need to render the utterances infelicitous, or unlicensed. The analysis can, again, be improved by requiring the agent to refer to a specific 'individual concept', such that he has the information that whatever is the value of that individual concept has the property denoted by W, cq. that it whistles. As a matter of fact, our analysis will come close to this one. Notice, however, that this notion of licensing essentially includes reference to individual concepts which an agent may or may not have in mind. For this reason it is not clear how such an analysis of licensing an utterance ϕ relates to the meaning of ϕ, standardly conceived.

An alternative statement of what licenses the utterance of open propositions we may seek to find in dynamic systems of interpretation. Dynamic systems of interpretation make substantial sense of the (linguistic) context dependent interpretation of variables or pronouns and, therefore, we can expect them to accommodate a sensible notion of the licensed uttering of open propositions. Indeed, GSV suggest that it is a dynamic semantic notion of support which is relevant here, although they provide us with no clues as to what that notion precisely amounts to. Two dynamic notions of support which can be found in the literature naturally suggest themselves, but none of the two seems to be adequate.

The notion of support which has been used to define entailment in systems of dynamic semantics is a too strong notion. A formula Gx is supported by an information state s if in s only valuations of x are considered possible which map x into an individual which is a G. This seems to be a sensible qualification of the universal requirement considered above, but it is still too strong a requirement. After a licensed utterance of $\exists xFx$, an utterance of Gx would only be licensed within a dynamic semantic system of information update, if one were to have the information that $\forall x(Fx \rightarrow Gx)$. But this is too strong. Consider the following example:

(12) A member of parliament visited the queen yesterday.
(13) He was dead drunk.

An agent may utter these two sentences if he has the information that the queen was visited by a member of parliament yesterday, who was dead drunk. It does not require him to know that *every* member who visited the queen yesterday was dead drunk. However, one would be required to have this kind of universal information for the utterance of the second sentence if the notion of a licensed utterance was stated in terms of (strong) dynamic support.

An alternative notion of weak entailment corresponds to (Chierchia 1992)'s weak notion of implication. Adopting such a notion, an utterance of Gx can be said to be licensed after a licensed utterance of $\exists xFx$, if one were to have the information that $\exists x(Fx \wedge Gx)$. Although this seems to provide a sensible qualification of the mere existential closure of Gx, it still appears to be too weak. Consider A's reply to B's question about a man A started to talk about:

(14) A: Yesterday, a man came into my office who inquired after the secretary's office. ...
(15) B: Was he wearing a purple jogging suit?
(16) A: If it was Arnold, he was, and if it was somebody else, he was not.

According to the weak entailment notion of licensing, A might motivate his reply by saying that several men came into his office yesterday, and that one, Arnold, was wearing a purple jogging suit, and that the others were not. But clearly it would be odd for him to do so. For, if he had started a story about Arnold, he should simply have said that he was indeed wearing he a purple jogging suit, and if he had started a story about any one of the others, he should have said that he was not.

It seems that neither a classical (universal or existential) nor a dynamic (strong or weak) notion of support offers us an appropriate notion of licensing open propositions. Still there is one more option available. One can be said to express an open proposition that Fx because of the information one has about a certain subject, not a determinate entity, that it is F. Before we can make this notion of licensing more precise we first have to present the notion of information which we borrow from systems of dynamic semantics.

On Information Aggregates

What, we think, we need in a definition of first order licensing is a suitable notion of aboutness. We want to make some (formal) sense of the (colloquial) locution that people (ought to) know what they are talking about, without this requiring their information to correspond to singular propositions. Very useful devices have already been developed for this purpose. *DRT*'s discourse representation structures serve to picture first order information, and dynamic semantic information states can be taken to model the contents of these discourse representation structures.

Discourse representation structures code information concerning subjects referred to by means of 'discourse markers' (Kamp 1981; Kamp and Reyle 1993). In systems of dynamic semantics this kind of information is fleshed out using sets of variable assignments, viz., the sets of assignments of values to the variables or discourse markers which are possible given the information which agents have about them (Heim 1982; Zeevat 1989; Dekker 1996). We like to view the mentioned information structures as modeling the (first order) contents of the information states of epistemic agents, but also,mor e generally, as that of texts and utterances and whatever else can be seen to contain first order information. For this reason we give them a general label "information aggregates".

Information aggregates are defined relative to a set of possible worlds W, a domain of individuals D and a set of variables V. (For the sake of simplicity, we assume that all worlds have the same domain.) Information aggregates contain information about 'subjects' which are referred to using these variables. Although it is not strictly necessary, we have chosen to use one distinguished variable v to refer to the world. For a finite set of individual variables $X \subseteq V$, with $v \notin X$, there is a domain of possible variable assignments $\{f \cup g \mid f \in W^{\{v\}}\ \&\ g \in D^X\}$, which we abbreviate as $D_W^{X_v}$. Information aggregates then are defined as follows:

Definition 1 (Information Aggregates): For finite $X \subseteq V$ and $v \notin X$
$\Sigma^X = \mathcal{P}(D_W^{X_v})$ is the set of information aggregates about X

The information that a subject x has a property $\mathcal{P}$ is captured by an aggregate σ such that each assignment $g \in \sigma$ assigns an individual d to x and a world w to v such that d has the property $\mathcal{P}$ in w. Similarly, the information that subjects x and z stand in the relation $\mathcal{R}$ is captured by means of sets of assignments σ such that for each $g \in \sigma$, the pair $\langle g(x), g(z)\rangle$ stands in the relation $\mathcal{R}$ in $g(v)$. Etc. (Notice that we use 'x', not only to refer to the variable x, but also to the possible values of x, i.e., the subject labeled by x. We trust this will not give rise to confusion.)

In what follows we will use $D(\sigma)$ to refer to the domain $D(\sigma)$ of σ, which is X iff $\sigma \in \Sigma^X$. We take it to be convenient to distinguish the absurd (empty) aggregates in Σ^X and Σ^Y for any two different sets X and Y, so one may assume that the various empty sets are labeled by their domains.

Information aggregates can be put to use in a system interpretation along the following lines, roughly. Definite and indefinite terms (descriptions, names, pronouns) can all be taken to set up or refer to subjects. Thus a sentence like (17) can be associated with an aggregate like (18) (with $X = \{x, z, u, w\}$):

(17) An actress gave it to him on a rainy Sunday.

(18) $\{g \in D_W^{X_v} \mid g(x)$ *is an actress in* $g(v)$ *&* $g(z)$ *is a rainy Sunday in* $g(v)$ *&*
$g(x)$ *gave* $g(u)$ *to* $g(w)$ *on* $g(z)$ *in* $g(v)\}$

This aggregate precisely contains the information about four subjects x, u, w and z that x is an actress who gave u to w on a rainy Sunday z. It will be clear that, for the purpose of modeling natural language interpretation, we need some device to code the information that the subjects u and w are pronominal and that they are in need of being resolved by the context. A sophisticated use of variables and a separate resolution relation help to do the trick. Consider the following example:

(19) A bear has escaped.

(20) It is chasing him.

Neglecting the second (demonstrative) pronoun 'him", the contents of (19) and (20) can be modeled by the aggregates (21) and (22), respectively:

(21) $\{g \in D_W^{\{x\}_v} \mid g(x)$ *is a bear in* $g(v)$ *& * $g(x)$ *has escaped in* $g(v)\}$
(22) $\{g \in D_W^{\{u\}_v} \mid g(u)$ *is chasing him in* $g(v)\}$

These two aggregates can be merged by resolving u into x, thus yielding the following product of information of the little discourse (19)–(20):

(23) $\{g \in D_W^{\{x\}_v} \mid g(x)$ *is a bear in* $g(v)$ *&* $g(x)$ *has escaped in* $g(v)$ *&* $g(x)$ *is chasing him in* $g(v)\}$

Thus the update potential of a sentence can be seen to be the sum total of its meaning (conceived of as an information aggregate) and a certain merging with other aggregates under some resolution relation. Below we present some more details of this merging operation; one may consult (Dekker 1997) for a more sophisticated treatment of anaphora along these lines.

Relating Information Aggregates

When we say that an information aggregate contains certain information about subjects referred to by $x_1, \ldots, x_n$ this means no more than that the aggregate has structured that information around n subjects. The fact that these subjects are labeled by the variables $x_1, \ldots, x_n$ is not of any great importance. An aggregate which contains the same information about subjects labeled $y_1, \ldots, y_n$ counts as an equivalent aggregate, like two discourse representation structures which are alphabetical variants of each other.

Given this, if we want to compare the information in any two aggregates, we first have to say which subjects in one are related to which subjects of the other. In our support relation we make this explicit. We will say that an information aggregate σ supports an aggregate τ under a linking relation r iff σ has more information about the subjects $x_1, \ldots, x_n$ in the domain $D(r)$ of r, then τ has about the corresponding subjects in the range $R(r)$ of r (where $D(r)$ is $\{x \mid \exists y\colon \langle x, y\rangle \in r\}$, and $R(r)$ is $\{y \mid \exists x\colon \langle x, y\rangle \in r\}$). The support relation can be defined as follows:

Definition 2 (Support): σ supports τ wrt r, $\sigma \trianglelefteq_r \tau$, iff $r \subseteq D(\sigma) \times D(\tau)$ and $\forall g \in \sigma \exists h \in \tau$: $g \equiv_r h$ (where $g \equiv_r h$ iff $g(v) = h(v)$ & $\forall \langle x, y\rangle \in r\colon\ g(x) = h(y)$

An aggregate σ supports τ wrt r iff r is a relation between the domains of σ and τ and every possible assignment in σ corresponds to one in τ which is identical modulo r. Thus, σ has more information about the subjects in the domain r than τ has about the subjects in the range of r.

The supports relation is quite a general one, because it may relate information aggregates relative to all different kinds of linking relations. Let us take a quick look at a couple of examples:

(24) if $r = \emptyset$, then $\sigma \trianglelefteq_r \tau$ iff $\{g(v) \mid g \in \sigma\} \subseteq \{h(v) \mid h \in \tau\}$, that is, iff σ has more information than τ in a classical sense

(25) if $r = \{\langle x, y\rangle\}$, then $\sigma \trianglelefteq_r \tau$ iff $\{\langle g(v), g(x)\rangle \mid g \in \sigma\} \subseteq \{\langle h(v), h(y)\rangle \mid h \in \tau\}$, that is, iff σ has more information about x than τ about y; so if y has some property $\mathcal{P}$ according to τ, then x has that property according to σ too

(26) if $r = \{\langle x, y\rangle, \langle z, w\rangle\}$, then $\sigma \trianglelefteq_r \tau$ iff σ has more information about x and z than τ about y and w ($\{\langle g(v), g(x), g(z)\rangle \mid g \in \sigma\} \subseteq \{\langle h(v), h(y), h(w)\rangle \mid h \in \tau\}$), so if y and w stand in some relation $\mathcal{R}$ according to τ, then x and z stand in that relation according to σ too

etcetera. It is relatively easily seen that $\trianglelefteq$ is reflexive under the identity relation, and transitive modulo composition of links: $\sigma \trianglelefteq_r \sigma$, for $r \subseteq \{\langle x, x\rangle \mid x \in D(\sigma)\}$; and if $\rho \trianglelefteq_r \sigma$ and $\sigma \trianglelefteq_s \tau$, then $\rho \trianglelefteq_t \tau$, for $t = s \circ r$.

A most interesting type of support is one which holds between two aggregates σ and τ under a relation r having all τ's subjects in its range. It is special enough to give it a specific label.

Definition 3 (Substantiation): σ substantiates τ (wrt l), $\sigma \mid\!\!\lessdot_{(l)} \tau$, iff $\exists r$: $\sigma \trianglelefteq_r \tau$ with $R(r) = D(\tau)$ (and $r = l$)

(We want to refer, below, to the specific relation under which one aggregate substantiates the other, so that's why we have added a possible label 'l' for r.) If an aggregate σ substantiates τ, then for all subjects $y_1, \ldots, y_j$ in τ there are subjects $x_{i_1}, \ldots, x_{i_j}$ in σ such that σ contains more information about $x_{i_1}, \ldots, x_{i_j}$ then τ has about $y_1, \ldots, y_j$. Thus, σ can be said to contain at least as much information about at least as many subjects as τ. Substantiation is a reflexive and transitive relation, but it is not antisymmetric. This is as we want it to be. Two aggregates can be substantially equivalent without being identical, simply because they use different names for corresponding subjects.

We can also define the product of information contained in a set Π of information aggregates relative to some linking function ϱ. A linking (or resolution) function ϱ assigns to each aggregate π in Π a finite linking relation $\varrho(\pi) \subseteq V \times D(\pi)$ and the product of Π under ϱ is an aggregate with domain $D(R(\varrho)) = \{x \in V \mid \exists y \exists \pi \in \Pi\text{: } \langle x, y\rangle \in \varrho(\pi)\}$. It is defined as follows:

Definition 4 (Information Product):

$$\bigwedge_\varrho \Pi = \{g \in D_W^{D(R(\varrho))_v} \mid \forall \pi \in \Pi\ \exists h \in \pi\text{: } g \equiv_{\varrho(\pi)} h\}$$

(Notice that it would make sense to use indexed information aggregates, because we might want to use the information of one information aggregate relative to different linking relations. We will neglect this point in the sequel.) The product operation is appropriately related to our substantiation relation, because it gives us the unique fixed substance of the exponents. For, if $R(l) = D(R(\rho))$ and $\forall \pi \in \Pi$: $R(\varrho(\pi)) = D(\pi)$, we find:

(27) $\sigma \mid\!\!\lessdot_l \bigwedge_\varrho \Pi$ iff $\forall \pi \in \Pi$: $\sigma \mid\!\!\lessdot_{\pi(\varrho) \circ l} \pi$

This shows that $\bigwedge \Pi$ contains nothing but the information contained in each of the $\pi \in \Pi$, relative to an appropriate linking relation. For, as a corollary of 27 we find that:

(28) $\forall \pi \in \Pi$: $\bigwedge_\varrho \Pi \mid\!\!\lessdot_{\pi(\varrho)} \pi$

4 Linking Information Aggregates

Now we have defined our notions of information, support and substantiation, it is time to get back to our original question: when can an utterance be said to be licensed? The idea motivating our analysis is this. Information aggregates in the actual world hang together in a world wide web of meaningful relationships. Aggregates contain information about real individuals in virtue of factual (causal, perceptual) links with (individuals in) the actual world; they may also contain information about subjects of other aggregates in virtue of factual links of a conversational or intentional nature. The last kinds of links come about when the subjects emerging in a discourse or acts of conversation are substantiated by the information of the agent who is responsible for the discourse or conversation act.

For this reason we will say that an utterance is licensed by an aggregate σ iff it is substantiated relative to some linking relation l:

Definition 5 (Licensing): σ licenses ϕ under l iff $\sigma \mid\!\!<_l [\![\phi]\!]$

An agent with information aggregate σ is licensed to utter ϕ if the subjects in the interpretation of ϕ are rooted in subjects of σ. The subjects mentioned by an utterance of ϕ are required to correspond to subjects in σ which are at least dressed with properties in σ which are attributed to them by ϕ.

This notion of licensing allows us to observe that if all subjects which arise in discourse are linked to specific subjects in the information aggregates of the responsible agents, and if all other subjects arise from perception of individuals, then all subsisting subjects are eventually related to specific individuals which can be said to be the ultimate *sources* of these subjects. Moreover, if the relevant perceptions are veridical, and if all exchange of information is substantiated, then it can even be shown that all existing information about subjects truthfully applies to their respective sources, the real individuals which agents are concerned with. (We must refer to the full paper for the formal proof of this fact.)

The notion of a substantiated utterance thus being motivated, we can now try and apply it to the phenomena which we came across in the second and third section of this paper. Consider again our little discourse (19)–(20), with associated information aggregates (21) and (22):

(19) A bear has escaped.
(21) $\{g \in D_W^{\{x\}_v} \mid g(x)$ *is a bear in* $g(v)$ *&* $g(x)$ *has escaped in* $g(v)\}$
(20) It is chasing him.
(22) $\{g \in D_W^{\{u\}_v} \mid g(u)$ *is chasing him in* $g(v)\}$

An utterance of (19) by an agent a with information aggregate ρ is licensed if $\rho \mid\!\!<_l$ (21) under some specific relation l linking x to a subject of ρ. And he is licensed to utter (20) if $\rho \mid\!\!<_{l'}$ (22) under a specific relation l' linking u to a subject of ρ.

In the case of an utterance of example (19) we may think of it as a dynamic semantic (or just pragmatic) fact about indefinites that they are used to refer to subjects which are not mentioned before, or to subjects the identity of which is deemed irrelevant. In example (20), however, a pronoun is used which needs to be resolved. An utterance of (20) thus invites us to look for a target to resolve the pronoun into and an obvious candidate is, of course, the subject which licenses the utterance of the first example. Indeed this target motivates the resolution which gave us the information product (23) above.

In order to make this a bit clearer, consider the product $\bigwedge_\varrho\{(21),(22)\}$ of the information of (21) and (22) under a resolution of u into x. (That is, let $\varrho(21) = \{\langle x, x\rangle\}$, and $\varrho(22) = \{\langle x, u\rangle\}$.) Next, assume l and l' link x and u to the same subject of ρ. In that case we find:

(29) $\rho \mathrel{|\!\!<}_l \bigwedge_\varrho\{(21),(22)\}$ iff $\rho \mathrel{|\!\!<}_l (21)$ and $\rho \mathrel{|\!\!<}_{l'} (22)$

Here we see that the specific links under which an utterances of (19) and (20) are licensed constitute sufficient motivation, or justification for taking the information product of the corresponding aggregates under a resolution u into x. This is an interesting point about licensing, which can be generalized. Let's say $[\![\phi \wedge_\varrho \psi]\!] = \bigwedge_\varrho\{[\![\phi]\!], [\![\psi]\!]\}$. We then find that:

(30) $\phi \wedge_\varrho \psi$ is licensed under l *iff* ϕ is licensed under $\varrho(\pi) \circ l$ *and* ψ is licensed under $\varrho(\pi) \circ l$

provided that $R(l) = D(R(\varrho))$ and $R(\varrho([\![\phi]\!])) = D([\![\phi]\!])$ and $R(\varrho([\![\psi]\!])) = D([\![\psi]\!])$. The last example shows licensing to be compositional.

It may be noticed that our notion of substantiation is both substantially weaker than the strong dynamic notion of support, and stronger than the weak dynamic notion of support. Our analysis can thus be seen to do quite a bit better then these two alternatives with respect to the little discourses in (12)–(16).

Cross Speaker Anaphora

Subjects introduced by means of indefinites noun phrases are generally assumed to be linked to subjects of the interpreter herself. Although a speaker may use an indefinite noun phrase in order not to reveal the identity of the described source, there should be a definite subject in the speaker's information aggregate licensing the indefinite. But also these licensing subjects can themselves be assumed to be derived from, or licensed by, subjects heard about from other agents, or indeed to derive from observed existing individuals. Especially in this last case (but also in the preceding one), cross speaker anaphora can be seen to be licensed.

Consider again the dialogue consisting of A's utterance of (1) and B's reply with (2):

(1) A man is walking in the park.
(2) He whistles.

If we come across a dialogue like this, B apparently knows (or thinks she knows) whom A is reporting about. Both A and B are required to have a subject in mind substantiating the contents of their respective utterances, and for the pronoun uttered by B to be resolved, we may take it to be the subject B thinks A reported about when he used the indefinite "a man". Leaving a (hell) lot of details aside, B's utterance can be said to be licensed if his aggregate $\tau_B \mathrel{|\!\!<}_l$ *u whistles*, $\tau_B \mathrel{|\!\!<}_{l'}$ *z is the source of* $[\![(1)]\!]$*'s* x and $l(u) = l'(z)$. The pronoun "he" is thus seen to be intended to be coreferential with "a man", and the information product of the two sentences can be obtained by resolving (the variable associated with) "he" into (the one associated with) the indefinite "a man". In this way the two sentences together are seen to provide the same information as if they had been uttered by one speaker, viz., that there is a man who is walking in the park and who whistles.

It must be noiticed that in the case considered above of course neither A nor B commits him- or herself to the product of information that there is a man who is walking in the park and who whistles. It suffices for A to have the information that a man is walking in the park, who constitutes a subject for him. For B it suffices that he has the information that the individual or subject B thinks A reported about whistles.

This is particularly relevant for the examples (5) and (6), uttered by R and N, respectively:

(5) A man is sleeping over there on a park bench.
(6) It is not a man, it is a woman, and she is not asleep, she is just sunbathing. Besides, it is not a parkbench.

Apparently, the second utterance relates to the individual which the first can be taken to be about. Probably, R and N have laid their eyes on a certain individual lying in the park, and R's utterance is licensed because he thinks the individual is a man who sleeps on a parkbench, and N's utterance is licensed because she takes it to be a sunbathing woman lying on something which is not a parkbench. That is, R's aggregate includes a subject presenting the individual as a man who sleeps on a parkbench, and N's aggregate includes a subject she takes R to be reporting about, and which presents the individual as a sunbathing woman lying on something which is not a parkbench. Clearly, N intends her utterance to be about the same individual as R's utterance, so the pronoun "it" can be seen to be coreferential with the indefinite "a man". But although the information product obtained under such a resolution of the pronoun is inconsistent, clearly, neither R nor N are committed to the validity of inconsistent information here.

The examples we have considered sofar seem to involve demonstratively present individuals, or, at least the assumption of demonstratively present individuals. But cross-speaker anaphora is not restricted to such cases. Agents are very well able to make more sophisticated guesses as to whom a speaker is referring to. Consider again the following dialogue:

(11) H: A magistrate from Gotham village has confessed battering young girls.
N: They say he suspected them of sorcery. Do you know if more magistrates confessed?
H: I don't know.
N: Do you know who he is?
H: No idea, he preferred to remain anonymous.

Apparently, H and N agree to be talking about one and the same magistrate, perhaps one they read about in the newspapers. Assuming that the information they gathered from the stories they read concerns a definite magistrate, they each may have subject substantiating their utterances; furthermore, assuming the other one's subject to relate to the same individual, they may take their utterances to be coreferential. Thus, the dialogue in (11) can be construed as being about one individual, expressing it is a Gotham village magistrate who confessed battering young girls, which he suspected of sorcery etc.

Notice that nothing in dialogue (11) prohibits H and N from not *really* speaking about the same thing. For one thing, the information of H and N may derive from different newspapers which reported about confessions of different magistrates. For another, the whole thing may have been made up by an overzealous reporter. However, although H and N may be deceived, or although they may be wrong in equating their subjects in the way they do, the sum of information (and the lack of information) conveyed in the dialogue (11) still is appropriately understood as being about one subject.

The Licensing Link

The cases of cross-speaker anaphora considered here involve reference to the subjects which a speaker has in mind when making a certain utterance. Observe that this would not make much sense if we would not require a licensed utterance to be substantiated by a definite link. When a speaker sincerely introduces subjects, or refers back to them, he must have definite subjects of his own in mind on our account. Thus, there can be said to be a *factual*, link between the subjects of utterances and the subjects of information aggregates licensing them. Notice that if all that was required was that there be *some* such link, then it seems there would be no ground for locutions like 'the subjects talked about'.

We think that the requirement that definite links support utterances also helps in explaining a couple of discourse phenomena other than cross-speaker anaphora. Consider again an utterance of example (14), followed by B's question (15) and A's answer (31):

(14) A: Yesterday, a man came into my office who inquired after the secretary's office.
(15) B: Was he wearing a purple jogging suit?
(31) A: He was indeed.

Suppose two men visited A's office yesterday, who both inquired after the secretary's office, suppose one (Arnold) was, and one was not wearing a purple jogging suit, and suppose A is fully aware of this. It seems A's answer to B can be simply true then. But it seems hard to see why this would be the case, if A's utterance of (14) is not understood as being about a definite subject of A which he knows as a man wearing a purple jogging suit.

For the same reason A may start a conversation with (12), and correct a reaction by B if B misidentifies the subject A started talking about:

(12) A: The queen was visited by a member of parliament yesterday. ...
(13) B: (I heard.) He was dead drunk.
(32) A: No, no, he was not; not the one I mean. Anyway, ...

Clearly A might agree with B that the queen was visited by a drunk member of parliament, but he still may be right in pointing out that he wanted to talk about another MP, who was not drunk. A can simply deny he is talking about the drunk MP, or deny that the one he talks about was drunk. Notice that if licensing only required there to be *some* substantiating link, then B's utterance could be argued to force A to talk about the MP who was drunk, which is absurd.

The requirement of a factual link may also help to rule out a reply like (16) to B's question (15) in the situation considered above:

(14) A: Yesterday, a man came into my office who inquired after the secretary's office.
(15) B: Was he wearing a purple jogging suit?
(16) A: (Yes, and no.) If it was Arnold, he was, and if it was somebody else, he was not.

If A were licensed to utter (14) without having a particular man in mind, then it seems (16) would be the most informative answer he could give to B's question. But, as we argued above, such a reply would be pretty odd in the sketched circumstances. On our explanation it is odd because A must have had a definite man in mind and, in the given circumstance, he could (and should) have given either a positive or a negative reply.

Finally consider the following continuation of dialogue (14)–(15). Suppose A replies:

(33) A: I don't know.

and suppose A continues the story he started to tell, and ends up concluding:

(34) A: So, after all it *was* the man with purple jogging suit.

On this scenario, if one were to inquire why A had initially replied that he didn't know whether the man was wearing a purple jogging suit a couple of answers may be perfectly acceptable. However, one answer we think is particularly marked: that A at the moment of uttering (33) did not yet know whom he was talking about, as if he had not yet made up his mind which story to tell.

The phenomena considered in this subsection are all naturally accounted for if the subjects set up in a dialogue must be substantiated by specific subjects of the aggregates of the speakers. Although speakers may very well have only partial information about the identity of the real individuals at issue, they should not be confused about which of their own subjects they are referring to.

5 Conclusion

In this paper we have argued that the phenomenon of cross-speaker anaphora is governed by the very same principles which govern single speaker anaphora. Speakers are generally required to convey information about subjects which they themselves have information about, no matter whether these subjects were mentioned first or introduced by themselves, or by other agents. There apepars to be only one notable difference between single and cross-speaker anaphora. In the case of single speaker anaphora it is obvious of course that a speaker can refer to a subject of his own which he himself already has introduced to a discourse. In the case of cross-speaker anaphora this may be more difficult because the hearer must find reasons to match a speaker's subject with one of his own. The last type of linking may be harder to motivate (or substantiate), and this fact may serve as an explanation for the fact that cross-speaker anaphora is more constrained than single speaker anaphora.

Our analysis has made crucial use of the idea that, although definite noun phrases come with uniqueness or familiarity assumptions, and indefinite noun phrases with novelty assumptions, both types of noun phrases can be fruitfully interpreted as a kind of free variables. The meanings of sentences can thus be conceived of as functions from the possible values of these variables, to the proposition which the sentence express under such valuations, and this Lewisian conception of meaning is fleshed out more generally in terms of sets of variable assignments, to model the contents of both sentences and information states alike.

It is precisely this conception of an information aggregate which has allowed us to make sense of a notion of aboutness which does not necessarily invoke singular propositions. It doesn't involve a too restrictive (too definite) interpretation of indefinite noun phrases, but it does provide us with a definite enough notion of a subject which can be used to model speaker's reference, single and cross-speaker anaphora, and linking more in general. Our first order extension of Grice's quality maxims in terms of substantiation justifies the assumption that epistemic agents have information about, and talk about, one and the same thing, without this requiring them to have uniquely identifying qualitative information of the relevant individuals.

We do think that the subjects of 'live' information states come with something like a uniqueness assumption, viz., the assumption that there is a true answer to the question which individuals these subjects eventually stand in for. We also think that this assumption constitutes the main motivation for linking subjects. Upon the assumption that two subjects eventually relate to one and the same individual, it makes sense to equate information about one with information about the other. But, again, people can be deceived, of course, and a plausible theory of information exchange should be able to account for the fact that agents intend exchange of information about one particular subject, when there is no real thing corresponding to it. And indeed our analysis allows for the possibility that agents are simply wrong when they think they refer to an individual, or when they think they refer to the same individual as someone else.

In section 2 we saw that cross-speaker anaphora is also possible in cases other than those of the kind described sofar. GSV have mentioned multi-speaker monologue as a separate kind, but we think an analysis of multi-speaker monologue anaphora straightforwardly fits in with the analysis presented in this paper. This leaves us with the cases of modally qualified cross-speaker anaphora. We think that these should fit in with the use of modally qualified single speaker anaphora. Consider again the examples (1) and (9):

(1) A: A man is walking in the park.
(9) B: He must be wearing blue suede shoes then.

As we argued, B's utterance can be taken to refer explicitly to the information expressed in A's utterance, and state that if indeed a (or: the) man is walking in the park, then he is wearing blue suede shoes. This, strikingly resembles cases of modal subordination. Consider Roberts' (single speaker) sequence:

(35) A thief might break into the house.
(36) He would take the silver.

It seems sensible to interpret (36) here as referring to the proposition in the scope of the 'might'-operator in (35). The second sentence can be used to claim that, if the possible state of affairs mentioned by example (35) turns out to be actual, that is, if in fact a thief breaks into the house, then, this thief will take the silver.

So, one might argue, both (9) and (36) refer back to a previously discussed proposition, the first to one expressed by another agent, the second to one qualified by a modal operator. In both cases the proposition invokes a subject which can be referred back to under a suitable modal qualification (cf., 'must' and 'would', respectively). We think the two types of examples are therefore amenable to an essentially similar treatment, as, for instance, that of (Geurts 1995) and (Frank 1997). Unfortunately, space prohibits us to go into the required detail here.

We want to conclude with a more general observation about the (dynamic) semantics/pragmatics interface. In this paper we have dissected the meanings of sentences from the update of information they may provide under some resolution relation. We think that the acceptability of information expressed by utterances and the linking of subjects is best conceived of as a pragmatic phenomenon, since it is concerned with utterances rather than sentences and with information about the context of utterance and about the information of other speech participants. Still, this does not at all need to hamper a systematic study of the logic of information update and exchange, for we can just take the outcome of a resolution process for granted and work on the assumption that various bits of information do get accepted. Systems of dynamic or update semantics can thus be seen to be specific systems of interpretation in which such studies are carried out precisely on these assumptions.

References

Beaver, D.: 1995, *Presupposition and Assertion in Dynamic Semantics*, Ph.D. thesis, CCS, Edinburgh

Chierchia, G.: 1992, Anaphora and dynamic binding, *Linguistics and Philosophy* 15, 111–83

Dekker, P.: 1996, The values of variables in dynamic semantics, *Linguistics and Philosophy* 19(3), 211–257

Dekker, P.: 1997, On denoting descriptions, in P. Dekker, J. van der Does, and H. de Hoop (eds.), *Proceedings of the Groningen Workshop on Definites*, OTS, Utrecht, in preparation

Francez, N. and Berg, J.: 1994, *A multi-agent extension of DRT*, Technical Report LCL 94–9, Technion—Israel Institute of Technology

Frank, A.: 1997, *Context Dependence in Modal Constructions*, Ph.D. thesis, Universität Stuttgart, Stuttgart

Geurts, B.: 1995, *Presupposing*, Ph.D. thesis, Universität Osnabrück, Osnabrück

Groenendijk, J., Stokhof, M., and Veltman, F.: 1997, Coreference and modality in the context of multi-speaker dialogue, in H. Kamp and B. Partee (eds.), *Proceedings of the 1995 Workshop on Context Dependence*, IMS, Stuttgart and UFAL, Prague

Heim, I.: 1982, *The Semantics of Definite and Indefinite Noun Phrases*, Ph.D. thesis, University of Massachusetts, Amherst, Published in 1989 by Garland, New York.

Kamp, H.: 1981, A theory of truth and semantic representation, in J. Groenendijk, T. Janssen, and M. Stokhof (eds.), *Formal Methods in the Study of Language*, Mathematical Centre, Amsterdam, Reprinted in J. Groenendijk, T. Janssen, and M. Stokhof (eds.), *Truth, Interpretation and Information*. Foris, Dordrecht, 1984

Kamp, H. and Reyle, U.: 1993, *From Discourse to Logic*, Kluwer, Dordrecht

Zeevat, H.: 1989, A compositional approach to discourse representation theory, *Linguistics and Philosophy* 12, 95–131

Changing the Common Ground

Jelle Gerbrandy*

1 Introduction

Often, in formal models of dialogue, the notion of a 'common ground' plays an important role: a body of public information which changes during the course of a conversation and is used to keep track of what has happened in the conversation, delimits the range of possible further utterances and influences the interpretation of those utterances. The reader need only skim many of the other papers in this volume to see that the idea is very much alive.

How exactly this common ground should be characterized is not agreed upon. To give some early examples: Lewis (1979) uses the metaphor of a 'conversational scoreboard' on which the relevant information about the 'moves' in the dialogue game are noted. Stalnaker (1978) speaks about 'presuppositions' as that 'what is taken by the speaker to be the common ground of the participants in the conversation, what is treated as their common knowledge or mutual knowledge.' Hamblin (1971) uses the metaphor of a 'commitment slate.' Yet other writers identify the common ground with 'that what is mutually believed.' Clark and Marshall (1981), for example, argue that it is necessary for a successful use of a definite description that it should be mutual knowledge what the definite refers to.

One can model changes in the common ground in one of two ways. In the first kind of model, a representation of the common ground is taken as primary, and the effect of utterances in a dialogue is modeled by showing how the utterance affects the common ground. The second approach starts out with the belief states of the dialogue participants considered separately. The effect of an utterance can then be modeled by the effect it has on the belief states of each of the separate agents. Since the contents of the common ground depend on the belief states of the participants, the effects on the states of the participants will also imply a change in the common ground; in that sense such models subsume a theory of how the common ground is changed.

Consider the following diagram:

* Department of Philosophy, University of Amsterdam. Parts of this paper were written at the CSLI of Stanford University; the Netherlands Organization for Scientific Reesearch (NWO) is gratefully thanked for sponsoring my visit there. I would also like to thank Henk Zeevat, Paul Dekker and Marco Hollenberg for their useful comments.

$$
\begin{tikzcd}
\text{Common Ground} \arrow[r, "\text{information change}"] & \text{Common Ground} \\
\text{Agents in States} \arrow[u] \arrow[r, "\text{information change}"] & \text{Agents in States} \arrow[u]
\end{tikzcd}
$$

At the bottom corners of the picture, there are agents which have certain information: each of them is in a certain information state. The arrow labeled 'information change' on the lower end of the picture represents the change in these states that is the effect of an utterance by one of the agents: this describes the second kind of model of the information change in dialogue. The first kind of model, where the effect of an utterance seen as a direct change of the common ground, is represented by the top part of the picture.

Of course, a piece of information can only play a role in dialogue if at least one of the participants in a dialogue is aware of that information. That makes it natural to assume that the common ground is determined by the information states of the agents: the vertical arrows represent some way of extracting the common ground from the agents' states.

The leading question in this paper is whether first taking the common ground in a certain model w of the states of the agents, and then changing the common ground according to some specified way, gives us the same result as first changing the model w, and then seeing what the common ground is in the result. Or, in other words: does the diagram above commute?

To make sense of this question, we need to be more precise about the filling in of the parameters: the kind of representation we use for states of agents and for the common ground, how these two are related, and what information change consists of. Of course, the answer to our question depends for a great deal on how we choose to fill in these parameters.

In the next two sections I will study the diagram using a classical possible worlds framework. In the first of these two sections (which is section 2) I discuss the relation between mutual belief and the common ground and discuss several definitions of mutual belief.

In the section after that, I show how a given function that describes information change can be 'lifted' to an operator that models 'mutual information change.' With this formal machinery, we have the tools to instantiate the informal picture above. We will look at operations of belief change such as expansion, contraction and revision. The main conclusions are negative: the diagram generally does not commute, not even for a relatively simple notion of belief change such as expansion. When considering weaker properties than commutativity, expansion fares fairly well, but I will argue that revision and contraction (and any kind of belief change operation that has certain minimal properties in common with these two) have properties that are incompatible with the assumption that the diagram behaves in a reasonable way.

In section 4, I will briefly study the same questions in a more general framework. The results will be similar to those of the preceding sections. The main purpose of this section is to show that the negative results hold for any kind of model that has certain minimal properties in common with the possible worlds approach.

The paper ends with a section entitled 'conclusions.'

2 Mutual belief and the common ground.

2.1 Possibilities

The standard way of modeling information of several agents in a possible worlds framework is by using Kripke models. Here, the information of an agent a in a world w is modeled by the set of worlds v that are accessible from w: those are the worlds that the agent a considers possible in w. We will adopt the same approach towards information, but use a different kind of model to implement it. The reason for not using Kripke semantics is that the possibilities defined below make it much more easy to define notions of information change.

Definition 2.1: Possibilities.
Let $\mathcal{A}$ be a finite set of agents and $\mathcal{P}$ a finite set of atomic sentences.

- Any function w on $\mathcal{A} \cup \mathcal{P}$ that assigns to each atomic sentence $p \in \mathcal{P}$ a truth value $w(p) \in \{0, 1\}$ and to each agent $a \in \mathcal{A}$ an information state $w(a)$ is a possibility.
- Any set of possibilities is an information state.

The intuition behind this definition is the following. We want a model of the world: atomic sentences are either true or false, and agents have certain information. Possibilities as defined here give us exactly that: to each atomic sentence, they assign a truth-value, and to each agent, they assign an object called an 'information state.' Secondly, we model the state of an agent in a traditional way: by the set of models of the world that are consistent with that agent's information. So, an information state will be modeled as a set of possibilities.

Assuming that our set of atomic sentences contains just a single sentence *It rains*, and we have only one agent, called Francisco, an example of a possibility is a function w such that $w(\textit{It rains}) = 1$ and $w(\textit{Francisco}) = \emptyset$.[1] So, in this rather depressing possibility it is raining, and Francisco's beliefs are not consistent: there is no possibility compatible with his beliefs.

Unfortunately, simply using standard set-theory as our background theory will not give us enough different possibilities to model everything we want to model. For example, there is no object in the ZFC set-theoretical universe that corresponds with a possibility w in which w itself is consistent with Francisco's information.

[1] In the following, we will often leave the precise structure of $\mathcal{A}$ and $\mathcal{P}$ implicit where it is not likely to lead to confusion.

This is why we use non-well-founded set theory, as it is developed in Aczel (1988). In this theory, the axiom of foundation is left out of the ZFC-theory, leaving us with an axiom system standardly denoted by ZFC^{-}. Instead, Aczel adds a new axiom, the axiom of anti-foundation, which for our purposes can be expressed as follows: "For each world in a Kripke model there is a unique possibility that is bisimilar to that world."[2]

This axiom guarantees us that we have enough possibilities to do epistemic logic: for every bisimulation class of Kripke models, there exists a corresponding possibility. If we want to use these models as a semantics for a modal language –and we do– this means that for every two Kripke models that are distinguishable in infinitary modal logic (modal logic with arbitrary conjunction added, a very strong language), there are corresponding possibilities that can be similarly distinguished. A forteriori, this holds for the weaker language we will use in our paper, which is a finitary modal language with an operator added for common belief.

Possibilities are studied in more detail in Gerbrandy (1997) and Gerbrandy and Groeneveld (1997). Possibilities are very similar to the states in the model for transition systems developed by Aczel (1988). Finally, the work in Barwise and Moss (1996) on using modal logic to describe non-well-founded sets is related to the present approach to epistemic logic.

Truth of classical modal sentences in a possibility can be defined in a way analogous to the definition of truth for Kripke models.

Definition 2.2: Let w be a possibility.

$$
\begin{array}{rcl}
w \models p & \text{iff} & w(p) = 1 \\
w \models \phi \wedge \psi & \text{iff} & w \models \phi \text{ and } w \models \psi \\
w \models \neg\phi & \text{iff} & w \not\models \phi \\
w \models \Box_a \phi & \text{iff} & \text{for all } v \in w(a) : v \models \phi
\end{array}
$$

There are two kinds of possibilities that we will be particularly interested in: the possibilities in which agents have introspective information (where they know exactly what information state they are in), and the possibilities in which their information is not only introspective, but also true. Introspection holds in those possibilities w such that $w(a)$, the state of a in w, contains only possibilities v in which a gets assigned the information state she is actually in, i.e. such that that $v(a) = w(a)$. Moreover, this property is also assumed to hold for each $v \in w(a)$. Truthfulness corresponds to the property that for each a, $w \in w(a)$: a considers the 'real world' possible. We will call possibilities that are both introspective and truthful 'S5-possibilities.'[3]

[2] The idea is that a Kripke model and a possibility are bisimilar just in case that for all practical purposes they have the same structure. Formally, a relation Z is a bisimulation iff its domain consists of worlds in Kripke models (i.e. pairs (K, x) where K is a Kripke model $(W, (R_a)_{a\in\mathcal{A}}, V)$, and x a world in K) and its range consists of possibilities. Moreover, if $(K, x)Zw$, then for each $p \in \mathcal{P}$, $V(x)(p) = 1$ iff $w(p) = 1$, and for each y such that $xR_a y$, there is a $v \in w(a)$ such that $(K, y)Zv$, and for each $v \in w(a)$, there is a y in K such that $xR_a y$ and $(K, y)Zv$.

We say that (K, x) and w are bisimilar iff there is a bisimulation Z such that $(K, x)Zw$.

[3] More formally, we take the class of introspective possibilities to be the largest class $\mathcal{I}$ such that for each $w \in \mathcal{I}$ and $v \in w(a)$: $v(a) = w(a)$ and $v \in \mathcal{I}$. The class of S5-possibilities is the largest class included

2.2 Mutual Belief

A modest part of the discourse on logic is concerned with the relation between mutual belief and beliefs of separate agents. By definition, a sentence ϕ is mutually believed iff each participant believes that ϕ to be the case, each participant believes that all other participants believe ϕ to be the case, etcetera, *ad infinitum*.

As I remarked in the introduction, one way of seeing the common ground is as seeing it as that which is mutually believed. This is the idea we will adopt in this section, so some comparison between this view and other views on the common ground are in order. First, I will argue that mutual belief can be seen as a stronger notion than that of the common ground – everything in the common ground must be (or can be taken to be) mutually believed – but that it depends on the view one adopts towards the common ground whether the converse holds, i.e. whether all mutual beliefs are in the common ground.

When seeing the common ground as the 'conversational scoreboard,' or as a 'commitment slate,' one can argue that whatever is on that scoreboard is independent of the beliefs of the separate agents. Lies, for example, will be added to the scoreboard in the same way as honestly believed utterances are (since the liar is committed to them in the same way as he is committed to honest utterances). Clearly then, there may be sentences on the scoreboard that are not believed, let alone mutually believed.

I think this point is valid, but it does not necessarily imply that the concept of mutual belief is irrelevant to the concept of the common ground when it is seen as a commitment slate. If we want a useful model of the common ground, it should also apply to conversations in which the participants try not to mislead. Given our little problem that is concerned with belief change resulting from utterances in dialogue, we can restrict the study of it to changes in the common ground that arise from honest utterances alone. In other words, we may assume that the participants *really* follow Grice's maxim of quality, and are really cooperative. Within this restriction on the kind of dialogue studied, it will never happen that sentences appear on the commitment slate or the conversational scoreboard that the participants are not committed to.

In any case, we will assume in the rest of this paper that the information in the common ground is in fact believed to be true by each of the participants.

A property of the common ground that, as far as I know, is shared in each model of the common ground that has anything to say about higher-order information (beliefs about beliefs and such) is that the common ground is in some sense 'publicly accessible': each of the participants knows what information is in the common ground. Given that whatever is in the common ground is believed by everybody, the public accessibility of the common ground implies that everybody believes that everybody believes the information in the common ground. We can repeat this argument to get arbitrary iterations of 'everybody believes ...'

So, under the assumption that the common ground is mutually accessible, and that all information contained in it is believed, it follows that all information in the common

in $\mathcal{I}$ such that for each possibility w in that class, $w \in w(a)$ for each a.

ground is mutually believed: there are at least as many things mutually believed as there are in the common ground.

The answer to the question whether all mutual beliefs are in the common ground depends on the view one takes of that common ground. If the common ground is seen as a kind of conversational scoreboard, or as only containing the information that the dialogue participants are committed to by utterances actually made, the answer will be 'no': surely many facts that are not explicitly stated in the dialogue can be taken to be mutual beliefs, such as the fact that the participants speak a certain language, that the speaker has an enormously big red nose, that there is a vase of flowers on the table between them, etcetera.

Other theories include all such information in the common ground. Clark and Marshall (1981), for example, argue that for a correct interpretation of definites such as the vase on this table,' facts such as that there is a vase on the table between the participants should be mutual belief. If one defines the common ground as 'all information that should be accessible for the dialogue participants so that their dialogue works,' the mutual belief that there is a vase on the table should be in the common ground as well.

2.3 Formal notions of mutual belief

Apart from the definition above, there have been many other definitions and characterizations of mutual belief. Jon Barwise (1989), in an article in one of the books on situation theory, compares three characterizations of the concept of common knowledge[4], and concludes that in situation theory, all three can be distinguished. In our format, these three notions can be, roughly, represented as follows:

The iterated approach is just a straightforward rewriting of the informal definition above:[5]

$$w \models C^{iter}\phi \quad \text{iff} \quad w \models \Box_{a_1} \dots \Box_{a_n}\phi \quad \text{for each sequence } a_1 \dots a_n \text{ of agents.}$$

The fixed point approach is based on the intuition that the mutual belief of a formula ϕ is a property (that is, a set) P of possibilities that holds of a possibility w just in case it holds in w that a knows that ϕ is the case, and each possibility that is in the information state of a also has the property P.

If we were to denote this property by '$\models C^{fix}\phi$', then this condition is formally expressed by the following equivalence:

$$w \models C^{fix}\phi \quad \text{iff} \quad \forall a \forall v \in w(a) : v \models \phi \text{ and } v \models C^{fix}\phi$$

[4] The term common knowledge and mutual belief are used interchangeably in the literature.

[5] The reason to use a 'C' to denote mutual belief is because the operator we will define is essentially the $C_{\mathcal{A}}$-operator of Fagin, Halpern, Moses, Vardi (1995). The reason they use this symbol is because they use the term 'common knowledge' instead of 'mutual belief.' Fagin et al. also study the logic of this operator, and provide a completeness theorem.

This does not uniquely identify a property though. For reasons explained in Barwise's article, we let $\models C^{fix}\phi$ be the *largest* property that satisfies the equation above.[6]

According to the 'shared situation' approach, a sentence ϕ is mutually believed just in case there is a situation σ in which (1) ϕ holds, and (2) the situation σ implies, or gives reason enough to assume, that each of the agents knows (or believes) that the situation σ in fact obtains, and (3) each of the agents does believe that σ obtains. This kind of definition has been proposed by Lewis (1969), Schiffer (1972) and Clark and Marshall (1981).

A typical example is a situation of sitting around a table on which stands a vase of flowers: such a situation would give each of the agents enough reason to assume that the fact that there is a vase of flowers on the table is mutual belief. Another typical example is the utterance of a sentence followed by an acknowledgment of the hearer: this situation would be reason enough to assume that the fact that the utterance is made is now mutually believed.[7]

If we identify a situation with a set of possibilities –'all maximal extensions' of that situation, if one wants, or 'all possibilities in which that situation obtains'– we can transpose Barwise's analysis in our framework and define:

$$\begin{array}{lll} w \models C^{share}\phi & \text{iff} & \text{there is a set of possibilities } \sigma \text{ such that:} \\ & & (1)\ v \in \sigma \Rightarrow v \models \phi \\ & & (2)\ v \in \sigma \Rightarrow v(a) \subseteq \sigma \text{ for each } a \\ & & (3)\ w(a) \subseteq \sigma \text{ for each } a \end{array}$$

If we compare the three definitions, it turns out that all three are equivalent:[8]

Fact 2.3: For each possibility w:

$$w \models C^{iter}\phi \ \Leftrightarrow\ w \models C^{fix}\phi \ \Leftrightarrow\ w \models C^{share}\phi$$

proof:

[From the iterated account to the fixed points] Assume $w \models C^{iter}\phi$. It is not hard to see that for each a and $v \in w(a)$: $v \models \phi$ and $v \models C^{iter}\phi$. Since we have defined $\models C^{fix}\phi$ as the largest set with exactly this property, it follows that $w \models C^{fix}\phi$.

[From fixed points to shared situations] Consider the set $\sigma = \{v \mid v \models \phi \text{ and } v \models C^{fix}\phi\}$. Then $v \in \sigma$ implies that $v(a) \subseteq \sigma$ by definition of C^{fix}, and clearly, $v \models \phi$ for each $v \in \sigma$. Assume $w \models C^{fix}\phi$. Then clearly, $w(a) \subseteq \sigma$, so $w \models C^{share}\phi$.

[6] Of course, the fact that such a largest set exists needs proof. We will omit it, just as we omit the motivation for choosing the largest property instead of, e.g. the smallest.

[7] Note that such knowledge is not meant to be infallible in any way. The negative results of Halpern and Moses (1991) in the context of message-passing systems shows that if one reads the 'knowledge' in 'common knowledge' in the strong sense as implying truth, it can never happen that any non-trivial information becomes common knowledge; at least not under the quite reasonable assumptions that message passing takes time, and is never completely reliable.

[8] Fagin et al. (1995) contains a proof of the equivalence of the iterated and the fixed point accounts

[From shared situations to the iterated approach] Assume $w \models C^{share}\phi$. We need to show that $w \models \Box_{a_1} \dots \Box_{a_n}\phi$ for each sequence $a_1 \dots a_n$ of agents. That this holds is easily proven by an induction on n. $\Box$

So, in a classical possible worlds framework (the definitions can be easily reformulated to apply to Kripke models, and the equivalence results will continue to hold) the three different characterizations of common knowledge collapse.

I am not sure whether this result should be seen as a positive or a negative one. In contrast to the analysis above, on Barwise's analysis of the three definitions in situation theory, all three definitions turn out to give different situation-theoretic notions of common knowledge. But the differences between the three kinds of definitions come up only at the transfinite level; restricting Barwise's analysis to models in which agents believe only finitely logically independent facts (a natural assumption to make on any agent), the three notions collapse also in situation theory. This makes it, at least to me, very hard to see how the distinctions between the three characterizations correspond to pre-situation-theoretic distinctions. To put it bluntly: it seems that situation-theory is making trouble where there was no trouble to be found.

Whatever the conclusion is, the fact that the three different characterizations come down to the same semantical characterization in our framework makes the choice between the definitions meaningless: we can take either one.

We will represent the common ground in a possibility w by an information state that contains exactly the information that is mutual belief. This information state contains all and only possibilities v for which it holds that one of the agents considers v possible (in w), or that one of the agents considers it possible that one of the agents considers v possible, etcetera. We let the notation $C(w)$ stand for this set of possibilities.

Definition 2.4: The common ground between the agents in a possibility w, $C(w)$, is the set of all possibilities v such that there is a sequence of possibilities and agents $w_0, a_0, w_1 \dots a_n, w_{n+1}$ such that $w_0 = w$, $w_{i+1} \in w_i(a_i)$ for each $i \leq n$, and $w_{n+1} = v$.

It turns out that this characterization is consistent with what we said previously: a sentence is accepted in the state $C(w)$[9] exactly when it is common knowledge in w:

Fact 2.5: $C(w) \models \phi$ iff $w \models C\phi$.

Before going back to our diagram, I would like to make some remarks about $C(w)$.

First, note that in $C(w)$, we have lost information about w: in general, there are w and v different from each other such that $C(w) = C(v)$. This also holds within the class of introspective possibilities. In particular, we cannot see from $C(w)$ alone where its possibilities 'come from': there is no way of telling from the structure of $C(w)$ whether some $v \in C(w)$ is there because some a thought it possible, or because some a thought some b considered it possible. We will return to this observation later.

[9] We say that a sentence ϕ is accepted in a state σ just in case ϕ is true in each possibility in σ

Another remark concerns the complexity of $C(w)$: it contains possibilities in which information of agents is represented, the information they have about each other's information, etcetera. Often, in models of dialogue, the common ground is not taken to be that complex at all: sometimes it contains only world-information (information that can be expressed by non-modal sentences), and in general, higher-order information (information about information) is only represented up to some very restricted finite depth. Also this point will be taken up in the next section, where we really start proving things about our diagram.

We end this section by noting some formal properties of common grounds, and comparing the common grounds introduced here with those of Zeevat (this volume).

Consider the following operation on sets of possibilities that collects all worlds considered possible in one of the possibilities of that set. We call the operation E.[10]

Definition 2.6: $E(\sigma) = \bigcup\{w(a) \mid w \in \sigma, a \in \mathcal{A}\}$.

If σ is a singleton set $\{w\}$, we will write $E(w)$ for $E(\{w\})$.

Fact 2.7: $C(w)$ is the smallest set σ containing $E(w)$ such that $E(\sigma) \subseteq \sigma$.

We end this section by comparing our common grounds to those of Zeevat (this volume). In his article, an information state σ is a common ground just in case it has the following property:

$$\sigma = E(\sigma)$$

Let's call this property the 'Zeevat property.' It turns out that many, but not all, possibilities have a common ground with the Zeevat property:

Fact 2.8: It holds that $C(w)$ has the Zeevat property iff $E(w) \subseteq E(C(w))$

That $C(w) = E(C(w))$ is not a very strong property of common grounds. For example, it is implied by introspection:

Fact 2.9: If w is introspective, then $E(w) \subseteq E(C(w))$.

Which means that each $C(w)$ belonging to an introspective possibility has the Zeevat property.

3 Changing the common ground

Suppose we are given an operator F over information states that expresses some sort of information change. What I have in mind is an operator such as 'expand with p' or 'revise with ϕ,' (Alchourrón, Gärdenfors, Makinson, 1985) or the update functions from update semantics (Veltman, 1996). The first question that I will try to answer here is what it means for a group of agents to apply such a function together; the second is how such functions behave in the diagram.

[10] The 'E' is from 'everyone.' The reason for this is that just as $C(w) \models \phi$ iff $w \models C\phi$, so it holds that $E(\{w\}) \models \phi$ iff $w \models \Box_a\phi$ for each a, i.e. just in case 'everyone knows ϕ.'

3.1 Multi-agent Expansion

We will start our discussion with one special and relatively simple case: that of expansion. Expansion with a certain sentence means simply adding the information contained in that sentence to the information you already have: in our case, that means discarding all possibilities in which the sentence is false:[11]

Definition 3.1: If σ is an information state, then $\sigma + \phi = \{v \in \sigma \mid w \models \phi\}$.

This definition may be familiar from update semantics, and if one takes classical logic as the 'base logic' in the work on belief revision (e.g. Alchourrón et al., 1985), this is essentially the definition of expansion used there. To keep things from getting too complicated, we will restrict our language to non-modal sentences in the following.

We are looking for a definition of 'mutual update' on the level of possibilities that corresponds with a change in the common ground. Consider the following definition, in which the notation $+^* \phi$ stands for a mutual expansion with ϕ:

Definition 3.2: $w +^* \phi = v$ iff $w[\mathcal{A}]v$ and for each $a \in \mathcal{A}$, $v(a) = \{v +^* \phi \mid v \in w(a) + \phi\}$.

In this definition, the notation $w[\mathcal{A}]v$ stands for the fact that w and v differ at most in the states they assign to agents in $\mathcal{A}$: w and v assign the same truth-values to the atomic sentences. For later use, we define the mutual update of an information state as $\sigma +^* \phi = \{v +^* \phi \mid v \in \sigma\}$. We will use this operation to change the common ground: it corresponds with learning that ϕ, and learning that all agents have learned that ϕ.

This definition is circular, but it does in fact define a unique function over possibilities.[12] The idea behind the definition is this: one of the participants a learning that all participants have expanded with ϕ is the same thing as a learning ϕ herself, and moreover, changing each of the possibilities in her resulting state to the effect that the participants have mutually learned that ϕ. The proof of fact 3.4 contains an example.

One way of viewing the operation $+^*$ is that it models a certain fact becoming common knowledge:

Fact 3.3: $w +^* \phi \models C\phi$.

Now that we have given all parts of our diagram a formal interpretation, we can redraw it:

[11] Often in the work on belief revision and expansion, information is represented by sets of sentences closed under some 'base logic.' We use classical possible worlds. However, if we assume this base logic to contain classical logic, the two modes of representation are equivalent.

[12] More precisely, the definition can be read as defining a system of equations (where objects of the form '$w +^* \phi$' are seen as the indeterminates), which has a unique solution by the axioms of non-well-founded set theory. In Gerbrandy and Groeneveld (to appear), a proof is given.

$$
\begin{tikzcd}
\sigma \arrow[r, "+^*"] & \tau \\
w \arrow[u, "C"] \arrow[r, "+^*"] & v \arrow[u, "C"]
\end{tikzcd}
$$

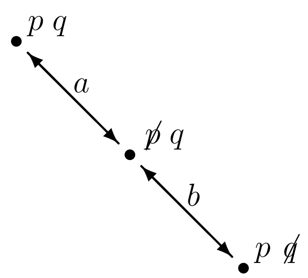
This diagram commutes just in case $C(w +^* \phi) = C(w) +^* \phi$. It turns out that this is not the case:

Fact 3.4: There are w and ϕ such that:

$$C(w +^* \phi) \neq C(w) +^* \phi$$

There is even an S5-counterexample.

proof: Consider the three possibilities given by the following equations:
$w_0(p) = 1$, $w_0(q) = 1$, $w_0(a) = \{w_0, w_1\}$, $w_0(b) = \{w_0\}$.
$w_1(p) = 0$, $w_1(q) = 1$, $w_1(a) = \{w_0, w_1\}$, $w_1(b) = \{w_1, w_2\}$.
$w_2(p) = 1$, $w_1(q) = 0$, $w_2(a) = \{w_2\}$, $w_2(b) = \{w_1, w_2\}$.
We can draw this model as follows:



In this picture, the topmost dot represents w_0, the middle represents w_1, and the lowest dot is w_2. I have not drawn reflexive arrows: but there should be both a and b-arrows going from each world to itself.

Consider $w_0 +^* p$. The state $w_0(a)$ contains two worlds, w_0 itself, in which p is true, and w_1, in which p is false, so $w_0(a) + p = \{w_0\}$. Applying the definition of $+^*$, this gives us that $(w_0 +^* p)(a) = \{w_0 +^* p\}$.

The state $w_0(b)$ contains only w_0 itself, so $(w_0 +^* p)(b) = \{w_0 +^* p\}$.

This means that $w_0 +^* p$ is that possibility in which both p and q are true, and each agent is fully informed about the world. We could draw the possibility by a single dot with a reflexive a and b-arrow. The common ground in $w_0 +^* p$ consists of just a single world: $C(w_0 +^* p) = \{w_0 +^* p\}$.

Consider now the common ground in w_0: $C(w_0) = \{w_0, w_1, w_2\}$. That means that $C(w_0) +^* p = \{w_0 +^* p, w_2 +^* p\}$. Since in $w_0 +^* p$, q is true, and in $w_2 +^* p$, q is false, $w_2 +^* p$ is different from $w_0 +^* p$, and hence, $C(w_0) +^* p$ is different from $C(w_0 +^* p)$. □

I could have chosen a counterexample that is less complex, but w_0 is the most simple example I could find that both is S5 and in which the update with p makes sense as the effect of an utterance in a dialogue between b and a. The possibility w_0 has the property $w_0 \models \Diamond_a \neg p$ and $w_0 \models \Box_b p$, which makes $w_0 +^* p$ a candidate for the result of b uttering 'p' in w_0. Since b believes p in w_0, she is in a position to say that p, and a has no reason to disagree: he considers p possible.

If one inspects w_0 in the proof of fact 3.4, one sees that a believes that b considers a $\neg q$-world possible only if $\neg p$ is the case. So, in a sense, the fact that $\neg q$ is a possibility in the common ground in w_0 depends on the fact that a considers $\neg p$ possible. In fact, a and b both know this, although it is not common ground that they both do. This is the reason that when one considers the mutual expansion with p, $\neg q$ disappears from the mutual beliefs, but that this is not reflected in the p-expansion of the common ground.

We get the same kind of result when we choose to represent the common ground in a less detailed way – containing only information about atomic sentences, for example.

Definition 3.5: σ contains less world-information than τ, $\sigma \preceq \tau$ iff for all $v \in \tau$ there is a $w \in \sigma$ such that $w[\mathcal{A}]v$

σ and τ are atomically equivalent, $\sigma \approx \tau$ iff $\sigma \preceq \tau$ and $\tau \preceq \sigma$.

Commutativity modulo atomic equivalence (which is in essence the same thing as representing the common ground as a state containing only information about atomic sentences), fails in the same way as it did before:

Fact 3.6: $C(w +^* \phi) \not\approx C(w) +^* \phi$

proof: Use the same counterexample as before. $\Box$

As I remarked above, mutual belief may be too strong a notion to use it as a model for the common ground. So one may view the property that the expansion of the common ground will never give you any results that are not also mutually believed in the mutually expanded possibility as a minimal correctness condition. In that respect $+^*$ behaves correctly:

Fact 3.7: $C(w +^* \phi) \subseteq C(w) +^* \phi$.

It turns out that the class of possibilities for which the diagram commutes, modulo $\approx$, coincides exactly with the class of possibilities where $E(w) = C(w)$: the class of possibilities in which 'everybody believes ϕ' implies 'it is mutually believed that ϕ:'

Fact 3.8: $E(w) \approx C(w)$ iff $C(w +^* \phi) \approx C(w) +^* \phi$ for all ϕ.

Because the language we are considering is not very rich in expressive power, we cannot prove a result to corresponding to the fact above with $\approx$ interchanged with real identity. We do have the following result:

Fact 3.9: If $C(w) = E(w)$, then $C(w +^* \phi) = C(w) +^* \phi$.

One can see the fact that the diagram commutes only for possibilities in which everybody's belief is mutual belief as a kind of diagnosis of the problem: in $C(w)$, the difference between first-order and higher-order knowledge has disappeared, and the only possibilities in which this does not lead to loss of information are those in which this distinction was not made in the first place.

3.2 Other Operators

The trick above, lifting an operator like expansion to a different operator corresponding to a mutual application can easily be generalized: we simply copy the definition for mutual expansion and apply it to an arbitrary function on information states.

Definition 3.10: Let F be any operation on information states. F^* is the following function over possibilities:

$$F^*(w) = v \text{ iff } v[\mathcal{B}]w \text{ and } v(a) = \{F^*(u) \mid u \in F(w(a))\}$$

Lifting an operation F to F^* has the following effect: each of the agents applies the operation F to his or her own information state, and then updates the possibilities in the resulting state with F^*. Just as in the case of mutual expansion, we will omit the proof of the existence and uniqueness of the function F^*.

I will show that assuming that $F^*(C(w)) \preceq C(F^*(w))$ for each w is inconsistent with assuming that F satisfies the postulates for either contraction or revision, together with the assumption that F is *flat*:

Definition 3.11: F is flat iff for all s and t: if $s \approx t$, then $F(s) \approx F(t)$.

An operator is flat just in case when the results of applying F to any two states that contain the same world-information will result in states that contain the same world-information too. I think that in general this assumption is not warranted, but when one assumes that F is meant to describe change in world-information only, the assumption is reasonable: if F expresses change in world-information only, its effects on the world-information should depend on world-information only.

It turns out when F is flat, monotony of F over $\preceq$ is a necessary condition for the property that $F^*(C(w)) \preceq C(F^*(w))$

Definition 3.12: An update operator F is monotone over an ordering $\preceq$ iff it holds that $\sigma \preceq \tau$ implies that $F(\sigma) \preceq F(\tau)$. We say that F is propositionally monotone iff it is monotone over $\preceq$.

Fact 3.13: If F is flat, and for each w, $F^*(C(w)) \preceq C(F^*(w))$, then F is monotone over $\preceq$.

proof: Assume F is flat, and $F^*(C(w)) \preceq C(F^*(w))$ for each w. Take any σ and τ such that $\tau \preceq \sigma$. Take any $v \in \tau$ (assuming that τ is not empty: in that case, $\tau = \sigma$, and we are finished), and define an S5-possibility w as follows:

$w(p) = v(p)$ for each $p \in \mathcal{P}$.
$w(a) = \{v \mid \exists u \in \sigma : u \approx v \text{ and } v(c) = w(a) \text{ for all } c \in \mathcal{B}\} \cup \{w\}$
$w(b) = \{v \mid \exists u \in \tau : u \approx v \text{ and } v(c) = w(b) \text{ for all } c \in \mathcal{B}\} \cup \{w\}$

Since $w(a)$, and each information state occurring anywhere in $w(a)$, is atomically equivalent to σ, and $w(b)$ and each information state occurring in $w(b)$ is atomically equivalent to τ, it follows that $C(w) \approx \sigma \cup \tau \approx \tau$. Since F is flat, it follows that $F(\tau) = F(C(w))$, and hence that $F^*(\tau) \approx F^*(C(w))$.

By assumption, we know that $F^*(C(w)) \preceq C(F^*(w))$.

By definition of C, $F^*(w)(a) \subseteq C(F^*(w))$, so $C(F^*(w)) \preceq F^*(w)(a)$. By definition of F^*, $F^*(w)(a) = F^*(w(a))$ and F^*, $F^*(w(a)) \approx F(w(a))$. Since we have defined w in such a way that $w(a) \approx \sigma$, we have by flatness that $F(w(a)) \approx F(\sigma)$.

Putting all of this together, we get that $F(\tau) \preceq F(\sigma)$. Since we chose σ and τ arbitrarily, we may conclude that F is monotone over $\preceq$. □

This result is interesting, because the notions of revision and contraction are not propositionally monotone. Consider for example the following two postulates that have been proposed as conditions on any contraction function:

K–3 If $\sigma \not\models \phi$, then $\sigma - \phi = \sigma$ (vacuity)

K–4 If $\not\models \phi$, then $\sigma - \phi \not\models \phi$ (success)

Fact 3.14: If p is atomic, then any function $-p$ that satisfies (K–3) and (K–4) is not propositionally monotone.

The proofs of this fact and the following are similar to those given in section 4.

To show that revision functions are not monotone over $\preceq$ either, we need the following two postulates:

K*3 If $\sigma \not\models \neg\phi$, then $\sigma * \phi = \sigma + \phi$.

K*4 If $\not\models \neg\phi$, then $\sigma * \phi \neq \emptyset$

Fact 3.15: If p is atomic, and $*p$ satisfies to (K*3) and (K*4) then $*p$ is not monotone.

4 A general framework.

I have shown how to formalize our informal picture in possible worlds semantics. In this section, I will try to assume as little as possible about the structure of information states, the common ground, or the relation between simple and mutual updates, and try to see which assumptions were really essential for the result to go through.

We start with some minimal assumptions that we need for representing agents with certain information. First of all, assume there is a set of agents $\mathcal{A}$, and a set of (information-)states $\mathcal{S}$ that those agents may be in. We will assume $\mathcal{A}$ to be finite, lets say $\mathcal{A}$ =

$\{1, \ldots, n\}$, and we will use s_i as variables over states that agent i is 'in'. (States may be represented by sets of possible worlds, by sets of sentences, discourse representation structures, databases, situation-theoretical objects, anything that suits your fancy.) We also assume that there is a transitive and reflexive relation $\preceq$ on $\mathcal{S} \times \mathcal{S}$, similar to the one we defined in the previous section. The idea is that this relation expresses 'containing less information about the world', i.e. it is a measure of information that disregards information about the epistemic states of the agents. We will write that $s \approx s'$ iff $s \preceq s'$ and $s' \preceq s$.

We want to be able to talk about agents having certain information in common, so we need a notion of agents being in certain states together. The simplest way to do this is by representing such a situation by a sequence $\bar{s} = \langle s_0 \ldots s_n \rangle$.

Another thing that we assume is that there is some function that extracts the common ground in a situation $\bar{s}$, and we assume that the common ground can be represented by the same kind of object that represent the states of the agents, i.e. we have a function C on situation $\bar{s}$, such that $C(\bar{s})$ is a state from $\mathcal{S}$. The following assumption can be seen as a minimal assumption on the function C:

common ground $C(\langle s_0 \ldots s_n \rangle) \preceq s_i$ for any $i \leq n$.

We assume that the common ground in a situation contains less information than each of the agents has in that situation. I don't think this is a controversial assumption in any way.

Now take an operation $F : \mathcal{S} \mapsto \mathcal{S}$ and a corresponding notion of a mutual application of this function $F^* : \bar{\mathcal{S}} \mapsto \bar{\mathcal{S}}$ that operates on sequences of states. I will propose a number of assumptions on these functions (all of which were assumed in the previous sections) which together are strong enough to give results similar to those we got in the previous section.

distributivity If $F^*(\langle s_0 \ldots s_n \rangle) = \langle t_0 \ldots t_n \rangle$, then $F(s_i) \approx t_i$ for all $i \leq n$.

To accept this postulate, keep in mind that $\preceq$ orders states with respect to world-information only. What the assumption says is that if the agents in $\mathcal{A}$ mutually perform the operation F, then their higher-order information may change in all kinds of ways, but the changes in the information they have about the world will be the same as when each of the agents would have applied the operation 'on her own.'

We need a third assumption to guarantee that we have enough states to work with:

fullness We assume that for every two states s and t such that $s \preceq t$, there is a situation $\bar{s}$ that contains a state t such that $t \approx t'$, and which is such that $C(\bar{s}) \approx s$

This is not a very strong assumption, I believe, but it may help to unravel the definition a little. Fullness says that for any two states s and t such that s contains less world-information than t, there is a situation $\bar{s}$ such that the world-information that is mutually known in $\bar{s}$ is the same as that contained in s, while one of the agents in $\bar{s}$ has the same world-information that is contained in t. [13]

The last assumption we make is the same as we did before:

flatness If $s \approx t$, then $F(s) \approx F(t)$.

Given these four assumptions, we can prove that if our diagram commutes, then F must be monotone over $\preceq$. In fact, we prove something slightly stronger, corresponding to fact 3.13, namely that monotony is a necessary condition for $F(C(\bar{s})) \preceq C(F^*(\bar{s}))$:

Fact 4.1: Assume that the four properties formulated above hold. Then is also holds that if $F(C(\bar{s})) \preceq C(F^*(\bar{s}))$, then F is monotone over $\preceq$.

proof: Take any s and t such that $t \preceq s$. Since $\mathcal{S}$ is full, we can find $\bar{s} = \langle s_0 \dots s_n \rangle$ such that $s \approx s_i$ for some $i \leq n$ and $C(\bar{s}) \approx t$. Since F is flat, $F(t) \approx F(C(\bar{s}))$. By assumption, $F(C(\bar{s})) \preceq C(F^*(\bar{s}))$.

Let $F^*(\bar{s}) = \langle t_0 \dots t_n \rangle$. By the assumption on the common ground, $C(F^*(\bar{s})) \preceq t_i$. By distributivity, $t_i \approx F(s_i)$, and using flatness again, we have that $F(s_i) \approx F(s)$.

Since we assumed that $\preceq$ is transitive, we can combine these observations and conclude that $F(t) \preceq F(s)$.[14] □

Since none of the operations considered above was originally defined to be applied to such abstract objects as the states introduced above, we still need to show that this abstract result applies to contraction or revision functions. Of course, we will not be able to prove anything about the original notions of expansion, contraction and revision. Instead, I will reformulate some postulates yet again (in general slightly weakening them), and then prove how failure of monotony follows from them.

To show that contraction functions are not monotone over $\preceq$, we need to reformulate the postulates for contraction in such a way that they apply to states in general. And for doing that, we need to extend our ontology: we need a language and a relation of $\models$ of 'acceptance' between $\mathcal{S}$ and this language. Think of $s \models \phi$ as meaning that ϕ is accepted in state s, that the information that ϕ is subsumed by the information in s, or that the information that ϕ is contained in s.

Consider the following postulates for contraction:

K'–2 $s - \phi \preceq s$.

K'–3 If $s \not\models \phi$, then $s - \phi = s$. (vacuity)

K'–4 If ϕ is not a tautology, then $\sigma - \phi \not\models \phi$. (success)

[13] Also for this assumption it is important to note that $\preceq$ pertains to world information only. Assuming this, I can see no reason for this assumption to fail in any of the representational frameworks that I know of. The proof of fact 3.13 contains a construction of such a state in a possible worlds model.

[14] I have skipped over matters pertaining to the possible partiality of the function F. If one defines monotony as a property that need only hold for values on which F is defined, the proof will work just as well.

The original formulation of (K'–2) uses a stronger notion of than $\preceq$, so the present formulation may be seen as a weaker version. (K'–3) is exactly the same as the original definition. (K'–4) introduces the notion of a 'tautology': this may be taken as a primitive notion, or it may be taken as defined as 'being accepted in each state' or as 'being accepted in the minimal state.'

To show a function $-\phi$ satisfying these three postulates is not monotone over ϕ, if ϕ is not a tautology, we need to be sure that $(\mathcal{S}, \preceq)$ contains enough structure.

We will assume that there are states s and t in $\mathcal{S}$, such that $s \not\models p$, $t \not\models p$, and for all u such that $s \preceq u$ and $t \preceq u$, $u \models p$. Moreover, we assume that there is in fact a u such that $s \preceq u$ and $t \preceq u$. (For an intuitively acceptable example, consider states s and t such that $s \models q$, $t \models q \to p$.)

We now have enough material to prove that $-p$ is not monotone over $\preceq$. For assume that $-p$ is monotone. We know that $s - p = s$ and $t - p = t$, by (K–3). Now take any u that contains more information than both s and t. By monotony, $u - p$ must contain more information than both s and t. But by assumption, every such state is one in which p is accepted, contradicting (K'–4).

The postulates for revision assume we have negation in our language, and that $\mathcal{S}$ contains an inconsistent state $\perp$. We will assume that if s is a state such that $s \models p$ and $s \models \neg p$ for some sentence p, then $s \approx \perp$. Consider the following postulates:

K*2 $s * \phi \models \phi$.

K'*3 If $s \not\models \neg\phi$, then $s \preceq s * \phi$.

K'*4 If $\neg\phi$ is not a tautology, then $\sigma * \phi \not\approx \perp$.

The postulate (K*2) is just the original postulate. It is not hard to see that (K'*3) is a weakening of (K*3), assuming at least that $s \preceq s + \phi$. Similarly, we have weakened (K*4) to the effect that if ϕ is not a contradiction, then a revision with ϕ will not be atomically equivalent with the inconsistent state.

Let p be such that $\neg p$ is not a tautology, and assume we have states s and t such that $s \not\models \neg p$ and $t \not\models \neg p$, and for all u such that $s, t \preceq u$, $u \models \neg p$. Assume moreover that there exists such a u. (Consider, e.g., $s \models q \to \neg p$, $t \models q$, similar as before.) Take any u such that $s, t \preceq u$. It holds, by (K'*3), that $s \preceq s * p$, and by monotony, that $s * p \preceq u * p$. Similarly, $t \preceq t * p \preceq u * p$. But then, by assumption, $u * p \models \neg p$. But according to (K*2), $u * p \models p$, from which it follows that $u \approx \perp$, which contradicts (K'*4).

5 Conclusions

In this paper, I have compared two ways to model changes in the common ground: changing a representation of the common ground directly versus seeing such changes as derived from changes in the belief states of the participants involved. It turns out that the two ways of modeling give different results for the two approaches.

The main conclusion to draw from the results of this paper is that mutual belief and common knowledge are not simply two sides of the same coin; at least not when one considers information change. This holds even for expansion.

Expanding the common ground may give results that are too weak when compared with mutual expansion, but they will not lead to additions in the common ground that are not also added when in the mutual expansion. If this discrepancy is a problem at all, I don't think it is a very serious one. Firstly because it seems that one does not seem to need all mutual beliefs to be in the common ground of a dialogue. But also because one of the reasons of using a separate representation of the common ground is that it is a less complicated way of modeling dialogue than keeping track of the states of the participants; this means loosing certain information about the relations between world-information and higher-order information, but fact 3.7 shows that this is basically harmless when considering expansion.

The result that a function that is flat has to be monotone for the changes in the common ground to be mutual beliefs, and that neither contraction nor revision are monotone seems more serious. One the other hand, both notions are notorious for their indeterminacy. What the results seem to say is that if one uses a simple deterministic function to model contraction or revision of the common ground, it may be that the resulting common ground will contain information that is not mutually believed. But if one takes a more lenient view on contraction or revision, as a process that involves some more or less arbitrary decisions on what kind of information to discard, i.e. if one considers the result of a revision process as, to a certain extent, unpredictable, it will be unclear in general what exactly is in the resulting common ground, and it will be even less clear to each of the participants what is mutually believed (since the latter involves reasoning about the belief change of the other agents, and their reasoning about each other's belief change, etcetera). The negative results seem to give just another argument that revision and contraction processes are not to be modeled by deterministic functions.

References

Aczel, P.: 1988, *Non-Well-Founded Sets*, Vol. 14 of *CSLI Lecture Notes*, CSLI Publications, Stanford

Alchourrón, C., Makinson, D., and Gärdenfors, P.: 1985, On the logic of theory change: partial meet functions for contraction and revision., *Journal of Symbolic Logic* 50, 510–530

Barwise, J.: 1989, On the model theory of common knowledge, in *The Situation in Logic*, pp 201–220, CSLI Lecture Notes, Stanford

Barwise, J. and Moss, L.: 1996, *Vicious Circles*, CSLI Publications, Stanford University

Clark, H. and Marshall, C. R.: 1981, Definite reference and mutual knowledge, in A. Joshi,

B. L. Webber, and I. A. Sag (eds.), *Elements of discourse understanding*, pp 0–00, Cambridge University Press, Cambridge, U.K.

Fagin, R., Halpern, J., Moses, Y., and Vardi, M.: 1995, *Reasoning about Knowledge*, MIT Press, Cambridge (Mass.)

Gerbrandy, J., *Dynamic Epistemic Logic*, To appear in the proceedings of the Second Conference on Information-Theoretic Approaches to Logic, Language, and Computation

Gerbrandy, J. and Groeneveld, W., *Reasoning about information change*, To appear in the *Journal of Logic, Language and Information*

Halpern, J. and Moses, Y.: 1990, Knowledge and common knowledge in a distributed environment, *Journal of the Association for Computing Machinery* 37(3), 549–587

Hamblin, C. L.: 1971, Mathematical models of dialogue, *Thoeria* 37, 130–155

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

Stalnaker, R.: 1978, Assertion, in P. Cole (ed.), *Pragmatics (Syntax and Semantics 9)*, pp 315–332, Academic Press, New York

Veltman, F.: 1996, Defaults in update semantics, *Journal of Philosophical Logic* 25, 221–26

Zeevat, H.: 1997, The common ground as a dialogue parameter, in *this volume*

Structural Mismatch in Dialogue*

Jonathan Ginzburg

Linguistics Programme, Dept. of English
Hebrew University of Jerusalem
Jerusalem, 91905
Israel
msjihad@pluto.mscc.huji.ac.il

1 Introduction

1.1 Some theses about Context and Conversation

The following are common working assumptions among workers on the semantics of dialogue—allowing for two variants, one for common ground–based approaches, the other for discourse structure–based approaches:

- **Equal Access to Context:** As a conversation proceeds a common ground emerges (discourse structure is constructed). A has her turn, reaches a transition relevance point (TRP); Then either A proceeds or B takes over *from the common ground point* at which A spoke (B tries to attach his contribution to the hitherto produced structure).

  (1) exemplifies why Equal Access seems a plausible assumption: A makes an initial utterance, a query, which either A or B can follow up on:

  (1) a. A(1): Who should we invite to the conference?
      b. A(2): Perhaps Noam, huh?
      c. B(2): Perhaps Noam, huh?

- **Dynamic Thesis:** Semantics is concerned with actions that change the common ground (with how a competent speaker should change her mental state).

- **Mutual belief construal:** The common ground represents the store of common knowledge/beliefs of A and B about *inter alia* the conversation.

* Thanks to Toni Benz and Gerhard Jaeger for the inspired idea and excellent organization of MunDial, which was a very pleasant and rewarding workshop. Thanks also to the participants for helpful comments and interaction. This paper summarizes a number of chapters from a (hopefully) forthcoming work tentatively titled *Questions and the Semantics of Interaction in Dialogue*. All are available by ftp from ftp.cogsci.ed.ac.uk:pub/ginzburg

1.2 The Turn-Taking Puzzle

In this section I present some data, which I will lump together as the Turn Taking Puzzle, and which I believe is problematic for the three theses articulated above, most directly for Equal Access.

(2) a. A: You're upset. Why? (unambiguously: 'Why is B upset')

b. A: You're upset. B: Why? (strong preference: 'Why does A claim B is upset'; 'why is B upset' weakly available)

c. A: You're upset. Why do I say that?

(3) a. A: Eric should be able to find a new job. Where? (unambiguously: 'where should Eric be able to find a new job')

b. A: Eric should be able to find a new job. B: Where? (Strong preference: 'where is A claiming Eric should be able to find a new job'; 'where should Eric be able to find a new job' available.)

(4) a. A: Ming will solve the problem. How? (Unambiguously: 'How will Ming solve the problem'.)

b. A: Ming will solve the problem. B: How? (Strong preference: 'How does A claim Ming will solve the problem'; 'How will Ming solve the problem' available.)

The data at issue with regard to (2-4) is the resolution of the bare factive–operator wh-phrases 'why', 'where' and 'how'. In the contexts described there are in principle two types of resolutions: one where the argument of the operator is the fact associated with the initial assertion ('the fact that B is upset'), the other where the argument is the fact associated with the initial utterance ('the fact that A asserted that B is upset'). The point these data illustrate most directly is that *who* utters a given form at a given point in the conversation can determine what interpretation that form receives. However, this context dependence is far stronger than indexicality ('I', 'you', 'now', 'here') since what is crucial is also who made the *previous* utterance. Notice that these data cannot be explained merely as a consequence of the differing coherence of an utterance depending on who makes the utterance: the resolution unavailable to A in (2) is coherent when it arises from a non-elliptical utterance, as in (2c). Rather, what this data seems to show is that which semantic objects are available to a particular dialogue participant, i.e. which entities she can exploit in elliptical or anaphoric resolution, depends on the role that participant has most recently played in the conversation.

Even more striking asymmetries are exemplified in a variant of the data in (2-4) in which the initial utterance is a query:

(5) a. A: Where was your Grandmother's sister born? Why? (Unambiguously: Why was she born *there*?)

b. A: Where was your Grandmother's sister born? Why? (Why do you ask where she was born?)

c. A: Where was your Grandmother's sister born? (and) Why am I asking this question?

(6) a. A: Who should easily be able to get a job (and) where? (Unambiguously: 'where can those easily able to get a job get a job?')

b. A: Who should easily be able to get a job B: Where? (Unambiguously: 'what place l is A asking about who should easily get a job at l?')

(7) a. A: Who solved the chess problem? (and) How? (Unambiguously: 'How did a solver solve the chess problem?')

b. A: Who solved the chess problem? B: How? (Unambiguously: 'What means of solution$_m$ is A asking who solved the problem by doing m?')

In (5-7) we see examples of dialogues where depending on who keeps the turn only distinct resolutions are possible:[1] the bare factive operators when uttered by A pick up on a fact that positively resolves the initial question A poses, whereas for B the resolution is to facts concerning A's initial utterance.[2] Once again, the differences cannot be explained by appealing to the pragmatic plausibility of the available reading, since as (5c) illustrates, the reading unavailable in (5a), makes for a perfectly coherent monologue if it arises from a non-elliptical form.

1.3 The structure of the paper

The Turn Taking Puzzle forces us to reconsider each of the assumptions articulated in 1.1: the data as such seems to directly refute Equal Access. More generally, it seems to suggest the need to study dialogue in a way that takes seriously the distinct roles dialogue participants play at different stages of a conversation. The aim of this paper is to sketch an approach that addresses this need. I will do this in a number of stages: I start by outlining a view of how dialogue participants (DP's) structure their dialogue gameboards (roughly: their versions of the common ground), what querying and assertion amount to, what coherence constraints the common ground imposes, and how this can be used to account for short answer ellipsis. This initial view will still be consistent with Equal Access, though will not *require* it. I will then suggest two modifications to this view: first, I will argue for a "localization" of the common ground structure of facts, tying this structure to the questions under discussion component. I will then consider one case which can lead to mismatches in the participants' dialogue gameboards, namely the unequal roles an utterer and addressee play at a particular point of the conversation. With this in hand I will return to reconsider and tentatively explain what is going on in the Turn Taking Puzzle.

[1] Actually, the data concerning B's options is ultimately more subtle: for instance if A and B are interrogating a third participant C, then B can also access the fact otherwise only available to A. For more discussion see Ginzburg 1997c.

[2] In fact, the resolution of B's is not straightforward to paraphrase—it is not obvious that a natural non-elliptical paraphrase exists.

2 The structure of context in dialogue

2.1 General Strategy

The notion of context required for dialogue semantics needs to steer a middle ground between individualism and sharing. On the one hand, if we are to have a chance of accounting for phenomena like the Turn Taking Puzzle, which indicate that at a given point in conversation the semantic options available to distinct participants can be distinct, our domain has to be the individual dialogue participant. Moreover, certain information needed to make sense of contributions is *unpublicised*: Crimmins 1993 and Asher 1993 argue for declarative attitude reports that their truth conditions depend on, in Crimmins' terms, agent internal *notions*, whereas I have argued in Ginzburg 1995a for interrogative attitude reports that their truth conditions are parametrized by agent-specific parameters such as goals and inferential capabilities. Much less controversial is the claim that the criteria for evaluating the coherence of a response needs to make reference to such agent-specific parameters. Thus, some of our rules regulating contextual change will need to be able to make reference to the DPs' individual mental states.

On the other hand, our view of context needs to allow for what seems like a basic fact of conversational interaction, namely the fact that DP's presuppose the existence of a common repository jointly built up and modified by the conversationalists as the conversation proceeds. Phenomena such as reference, in particular naming, and presupposition, for instance factivity, seem hard to explain without allowing for some such notion (*the common ground* (e.g. Stalnaker 1978), *conversational scoreboard* (Lewis 1979), or *discourse-structure* (Polanyi 1987, Asher 1993). Indeed, I will argue that certain important preparatory and coherence conditions for dialogue require stating relative to a common—ground-like construct.

In attempting to defuse the tension between individual and common aspects of "context", I will adopt the following strategy: I take the individual DP, more precisely her total mental state, as the basic domain of description. Each such mental state, nonetheless, will be taken to be partitioned in two: the first component is a *quasi*-shared object, each DP's version of the common ground; rather than using the latter term I will prefer to call it the DP's *dialogue-gameboard*, to suggest games like BATTLESHIPS where distinct individuals come to classify a single situation but along the way might have distinct representations. The second component of the DP's mental state is the *non-publicized* aspects of each participant's individual mental state, the DP's *unpublicized mental situation* (UNPUB-MS(DP)). More precisely, any information not explicitly considered to be part of the DGB is considered part of the DP's *unpublicized mental situation*
(DP | UNPUB-MS).

DP's MENTAL $\boxed{\text{DGB}}$ STATE

My main concern here, given the broadly semantic emphasis of the paper, will be on the nature of the DGB. Moreover, I will treat a DGB as a "semantic entity", that is, as individuated in terms of semantic objects such as facts, questions, and propositions. This move is based on the assumption that in conversational interaction DP's manipulate such semantic objects: an assertion is understood, roughly speaking, as exhorting the DP to accept a given item of information as a fact, a query as exhorting the DP to provide information about a certain question etc. Thus, rules of conversational interaction will involve operations on the DGB, though certain of these will involve UNPUB-MS parameters. This is consistent with theories of attitude reports like those advanced by Crimmins 1993, and Cooper and Ginzburg 1996, where e.g. belief reports are treated as relating an agent, a proposition p, and a mental situation ms, such that p is the content of ms.

2.2 The structure of the DGB

Following Ginzburg 1995b I assume that the DGB is structured by at least the following attributes:

- FACTS: set of commonly agreed upon facts.

  The conception of facts assumed here takes a fact to be a SOA made factual by (at least) one situation. FACTS, then, is a set of factual SOA's but in addition carries some structure. What structure? In standard situation theoretic practice, the set of SOA's constitutes a Heyting algebra, so in particular is closed under arbitrary meets and joins and also carries a dual operation. The set of FACTS, however, is obviously *not* closed under dualization; quite the contrary[3]. I assume, then, in common with Asher 1993, that FACTS is closed under meets and joins. The main ramification of this in practice is that, for instance, once an assertion **that p** is accepted and FACTS is incremented with **the fact that p**, and subsequently an assertion **that q** is accepted and FACTS is incremented with **the fact that q**, then both **the fact that p and q** and **the fact that p or q** are also included in FACTS.

- QUD ('questions under discussion') : a set that specifies the currently discussable questions, partially ordered by $\prec$ ('takes conversational precedence'). If q is maximal in QUD, it is permissible to provide any information specific to q using (optionally) a short-answer.

- LATEST-MOVE: content of *latest move* made: it is permissible to make whatever moves are available as reactions to the latest move. For current purposes, this attribute is rather straightforward: it holds of the information characterizing the content of the most recent dialogue move—**'A asserted that p'**, **'A asked q'** etc.

[3] Assuming the world is coherent and that SOA's are *persistent*, then $s \models \sigma$, implies $s \not\models \overline{\sigma}$.

3 Coordinating discussion: a protocol

3.1 Queries: initial considerations

Let us start by considering the simplest case: assume that the conditions for *posing* a question are satisfied and that a question q is indeed posed. The smoothest development is that q gets adopted by both parties and a discussion concerning it ensues. Querying would then involve the following contextual development:[4]

(8) cooperative querying

1. A poses q:
   - q becomes maximal in A | QUD: QUD:= QUD $+_{QUD}$ q
   - A | LATEST-MOVE: A ASK q.
2. B realizes a query was posed and accepts the question:
   - B | LATEST-MOVE: A ASK q.
   - q becomes maximal in B | QUD: QUD:= QUD $+_{QUD}$ q
3. B provides a response u that addresses q:

   u addresses q if and only if either:
   - content(u) = p, and p ABOUT q
   - content(u) = q_1, and q_1 INFLUENCES q

In (8) by the phrase 'q becomes maximal', notated QUD:= QUD $+_{QUD}q$, I mean that QUD undergoes the following update operation: q is added to the previous QUD-set; the ordering $\prec$ is modified in such a way that all the previous relations are retained but also for every element q_0 in the previous QUD-set: $q_0 \prec q$.[5] Thus:

(9) q becomes maximal in QUD: QUD:= QUD $+_{QUD}$ q

a. QUD:= QUD $\cup\{q\}$

b. $\prec := \prec \cup\{\langle q_0, q\rangle \mid q \in QUD\}$

Consider (10):

[4] Roberts 1995 develops an approach to querying and more generally the structure of context, which is similar in certain respects to the current proposal. For discussion and comparison see Ginzburg 1997b.

[5] Also: 'INFLUENCE' is the converse of the relation 'DEPEND', originally discussed in Karttunen 1977. Within the current framework, assumptions about this relation follow Ginzburg 1995a.

(10)

A(1): Who's coming tomorrow?
B(2): Several colleagues of mine.
A(3): I see.
Option 1: A(4'): Will Mike be coming?/ Is anyone else coming?/ Mike is coming too.
Option 2: A(4"): What is Merle doing these days?
Option 1': B(4'): Mike is coming too.
Option 2': B(4"): Are you planning to do anything much?

Here A poses the question in (1), B then addresses the question in (2), and his response is accepted in (3). At this point, the dialogue can take at least two directions: the original question can continue to be discussed by either participant (as in options 1/1') or a new question can be introduced (as in options 2/2'). Whether the former or the latter direction is pursued is simply *not* a matter that the DGB determines. Rather, this is an issue fixed at the level of UNPUB-MS's: if the accepted information resolves a given question q in QUD, relative to a DP's publicized or unpublicized information, this seems to constitute a sufficient condition for downdating q from QUD.[6] However, as discussed in detail in Ginzburg 1995a, the notion of goal–fulfilling information is more inclusive than resolving information. To put it differently, a querier can be satisfied with a response (or sequence of responses) without being thereby provided with resolving information. Or, in current terms, whether a DP decides that information she accepts warrants downdating a question from QUD is, ultimately, determined by her goals represented in her UNPUB-MS. Such downdating is, in case QUD does not become empty as a result, often accompanied by a "topic–changing"/"popping" cue word such as so or anyway. Thus, we have the following principle:

(11) **QUD DOWNDATING PRINCIPLE (QDP) [initial version]:** Assume q is currently maximal in A | QUD, and that there exists $\psi \in$ A | FACTS such that $\psi \Rightarrow_{UNPUB-MS(A)} goal-SOA(UNPUB-MS(A))$
Then, and only then permit DP to remove q from QUD:
A | QUD:= A | QUD $\setminus\{q\}$
$\prec := \prec \setminus \{\langle q_i, q_j\rangle \mid$ either $q_i =$ q or $q_j =$ q $\}$

3.2 Assertions: initial considerations

Before we can actually provide even a fairly schematic analysis of dialogues such as (10), we need to say something about assertion in dialogue, noting along the way a number of important differences between dialogue and text/monologue. Indeed, both the view of context incrementation deriving from Stalnaker and the discourse-structure tree-based view face certain problems. The crux of the matter is that when a new assertoric contribution is encountered, it cannot, as is the case in the various standard approaches to discourse semantics, be *attached* simpliciter or added into FACTS. Consider a situation in which

[6] In saying this, I mean to emphasize that not solely public information need be exploited in determining whether the question is resolved.

A makes an assertion *that p*. Now if FACTS is to serve as some sort of common ground repository, one which underwrites the felicity of presupposition-carrying utterances such as 'given that ...', 'Since we know that...' etc, A *cannot*, with an important caveat discussed in Ginzburg 1997b, update FACTS *before* receiving acceptance from B. In what way is A to modify his gameboard in the meantime? One clearly *monotonic* effect of asserting *that p* is to raise the issue of *whether p* for discussion: either there is immediate consensus about the question, in which case FACTS gets incremented with p, or the question gets discussed which can (but need not) lead to subsequent incrementation of FACTS with either p or $\neg p$. This is illustrated in the following dialogue: at the 3rd turn two options are illustrated—acceptance, (3"), and discussion, (3'). (4') continues discussion of the issue *whether Helen will agree to come*. (5') exemplifies the case where the discussion terminates with the original assertion being accepted, whereas (6) exemplifies the case where the negation of the orignal assertion wins out:

(12)

A(1): Who will agree to come?
B(2): Helen.
A(3'): I doubt Helen will want to come after last time. A(3"): I see
B(4'): Nah, I think she's forgiven and forgotten.
A(5'): OK. A(5"): No, she definitely has not.
B(6): OK.

I propose, therefore, that what an asserter does to her DGB immediately *after* asserting is to update her QUD with the question *whether p* as its topmost element. Let us assume, in line with the assumptions adopted sofar concerning querying, a relatively "cooperative" contextual development. The consequence of this is that B also adopts *whether p* (henceforth denoted as 'p?') as QUD–maximal: this step can but need not involve producing an utterance that expresses this question, such as 'really?'. I suggest that B now faces two generalized options: either to accept the assertion or discuss the issue. The former option is slightly more complex: it involves the following three steps:

1. **B makes a utterance u about p?**: u positively resolves p?.

2. **B increments her FACTS:**

   (i) B | FACTS: = FACTS $+_{fact-closure} SOA(p)$

   Here SOA(p) is the fact that must hold iff p is true. That is, if p = $(s!\sigma)$, then SOA(p) is $\langle\langle\models, s, \sigma\rangle\rangle$; $+_{fact-closure}$ is an operation which unions and closes under $\vee$ and $\wedge$—when no confusion can occur I omit the subscript 'fact-closure'.

3. **B downdates p? from QUD**: given that B's FACTS contains a fact that resolves p?, by the QUD Downdating Principle B can downdate p? from QUD.

The other option is simply not to accept the assertion and offer a contribution about p?: B produces an utterance specific to p?.

This protocol is summarized in (13):

(13) Cooperative Assertion

1. A asserts p:
   - p? becomes maximal in A | QUD.
   - A | LATEST-MOVE: A ASSERT p.
2. B's reaction:
   - B | LATEST-MOVE: A ASSERT p.
   - p? becomes maximal in B | QUD.
3. **Option 1: Acceptance:**
   - **B makes an *affirmative* utterance u about $p?$.**
   - **B increments her FACTS:** FACTS:= FACTS $+_{fact-closure}SOA(p)$
   - **B downdates p? from QUD**.
4. **Option 2: Discuss**:
   - B provides a response that addresses p?

Let us now consider the kind of DGB-based analysis we can offer for a dialogue, for instance one where the assertion leads to discussion:[7]

(14)

A(1): Who will agree to come?
B(2): Helen and Jelle.
A(3): I doubt Helen will want to come after last time.
B(4): Nah, I think she's forgiven and forgotten.
A(5): OK.

(15)

(1): A | QUD: q_1
(2): B | QUD:= q_1; asserts p_1 about q_1: B | QUD:= $q_1 \prec p_1$?
(3): A | QUD: $q_1 \prec p_1$?
asserts p_2 about p_1? : A | QUD:= $q_1 \prec p_1? \prec p_2$?
(4): B | QUD:= $q_1 \prec p_1? \prec p_2$?
asserts p_3 about p_2?: B | QUD:= $q_1 \prec p_1? \prec p_2? \prec p_3$?
(5): A | QUD: $q_1 \prec p_1? \prec p_2? \prec p_3$?
accepts p_3: A | FACTS:= soa(p_3);
downdates p_3? from QUD: A | QUD: $q_1 \prec p_1? \prec p_2$?
Given that p_3 resolves p_2?: downdates p_2? from QUD:
A | QUD: $q_1 \prec p_1$;

3.3 Dialogue-level Appropriateness conditions

With an initial proposal in place of how querying and assertion work in dialogue, we can turn to discuss the conditions in which such moves can be made. More specifically, I will

[7] An issue I have glossed over is how A reacts to B's acceptance. For discussion see Ginzburg 1997a.

be concerned with constraints that derive from and pertain to the DGB: given that a DP is a participant in a conversation, and given that the conversation has reached a certain configuration κ, in what way can this influence the DP's actions? The following condition attempts to combine a (necessary) condition requiring that the DGB *lack* resolving information about a given question with a (sufficient) condition that licenses introducing a question into the context if it has an "antecedent" in QUD. However, in order for this to have a chance of being at all successful some additional condition needs to be considered: one can certainly pose questions when QUD is *empty*, both at an initial stage of a dialogue and at transition points, where as far as a DP is concerned, the current issue can be downdated:

(16) Question Introduction Appropriateness Condition (QIAC):

Given a DGB configuration δ = DGB(A), a question q can be introduced into A | QUD by A when and only when

1. Either of the following 2 conditions applies
   (a) A | QUD is empty
   (b) Maximal in A | QUD is a question q_1 such that q influences q_1 relative to UNPUB-MS(A).
2. There does not exist a SOA τ such that $\tau \in$ A | FACTS and τ resolves q relative to UNPUB-MS(A)

Note that the condition is formulated not merely as one on queries but rather as a constraint about the introduction of questions into QUD, whether directly by querying or as a side effect of other dialogue acts. Thus, we have already seen how assertion *that p* can be analyzed as introducing a question *whether p* into QUD, and we shall see some more examples of this type of analysis in the sequel. The constraint will preclude a question from being introduced into QUD if information that resolves it is already established in the DGB. For the case of assertion, it will enforce the infelicity of asserting a proposition which *contradicts* a common ground fact, as well as that of a proposition which is already in the common ground.

3.4 Short Answers

Let us turn now to the first indubitably semantic application of the DGB, namely the interpretation of phrasal utterances. As we have already seen in various examples in this paper, one of the most obvious ways in which a query use of an interrogative i_0 changes the context is to enable elliptical followups that "syntactically match"[8] the interrogative phrase(s) of i_0. Indeed, there are various examples that suggest that elliptical contributions are possible, in principle, arbitrarily far away from the turn in which the question was posed as long as the question remains under discussion.

In this section I want to illustrate how such utterances can be interpreted by appealing to a rule that makes what is, I believe, non-eliminable reference to QUD.[9]

[8] In a sense made explicit in Ginzburg 1996.

[9] The rule proposed here presupposes a purely semantic approach to interpretation of short answers, in contrast to the common and perhaps common-sensical assumption that these involve *ellipsis* at some

For current illustrative purposes I will consider only short answers to unary wh-questions, though Ginzburg 1996 considers the general n-ary case, where a number of additional interesting issues arise. The rule in (17) says that in a DGB configuration DGB_0, any XP can be expanded as an S whose content is calculated as follows: it is a proposition of the form $(s!\sigma)$. Here s is the situation component of the question maximal in QUD; σ arises by predicating of the XP the abstract component of the question maximal in QUD.

(17) a. S $\rightarrow$ XP

b. Content(S)[DGB_0] = (SIT(DGB_0 | Max-QUD) ! λ-Abstr(DGB_0 | Max-QUD)[Content(XP)])

The main innovation here from the so-called categorial approach to interrogatives pertains to the reference to QUD, which connects up to context in an explicit way, in particular enabling an account of short-answers used an arbitrary distance away from the question to which they pertain. Thus, in (18(4)), 'Jelle' can be interpreted as 'We should invite Jelle for tomorrow' because at that point in the dialogue the question expressed by a use of 'who should we invite tomorrow' is QUD–maximal:

(18)

A(1): Who should we invite for tomorrow?
B(2): Who will agree to come?
A(3): Helen and Jelle and Fran and maybe Sunil.
B(4) : I see. So, Jelle.

(1): A | QUD: q_1
(2): B | QUD: q_1; asks q_2 such that q_2 influences q_1;
B | QUD: $q_1 \prec q_2$
(3): A | QUD:= $q_1 \prec q_2$; asserts p_1 about q_2:
A | QUD: $q_1 \prec q_2 \prec p_1$?
(4): B | QUD:= $q_1 \prec q_2 \prec p_1$?
B | FACTS : soa(p_1);
downdates p_1?, q_2 from QUD: B | QUD: q_1;
asserts p_3 about q_1:
B | QUD:= $q_1 \prec p_3$?

(19) a. q_1: $(s_{dinner}?\lambda x\langle INVITE-ABLE, x\rangle)$

b. Content of answer phrase: (reference to the individual) **j**.

c. Content of the short answer: $(s_{dinner}!\langle INVITE-ABLE, j\rangle)$

level of syntactic analysis. Ginzburg 1996 tackles this latter assumption head on: I argue there that the resolution process involved in short answers requires a new view of ellipsis resolution, one that remains semantically–based but does deviate from a purely semantic approach, by incorporating a notion of a *syntactic presupposition*. I argue there that a syntactic view of ellipsis resolution for such cases is unworkable from syntactic, semantic, and processing considerations.

4 Localizing the structure in FACTS

4.1 Structure and ellipsis

The structure we have imposed on FACTS hitherto was motivated by essentially presuppositional concerns: given that FACTS initially comprises a set F_0 of SOA's and that σ is to update FACTS, what SOA's need to be added to F_0 to represent those items of information that the conversational participants *take for granted*, as evinced by locutions such as 'given that ...', 'since we know ...' etc. However, this leads to an unattractive asymmetry with QUD: whereas the latter intermittently grows *and* declines, allowing us *inter alia* to construct a theory of resolution for short answers, FACTS only grows and has nothing that could be termed *local* structure. One problematic consequence of this concerns ellipsis involving factive operators. Consider (20): in (4) 'why' has two possible resolutions, given in (20b). However, if this dialogue continues from (3) with the turns (4')-(6'), the resolution possibilities change, in particular the facts previously available as resolutions are no longer available:

(20) A: Who's left recently?
B: Bill.
A: Is he the guy everybody hates?
B: Yeah.
A(4): Hmm. Do you know why?

a. why: why Bill left recently. or: why does everybody hate Bill.
A(4'): Uh huh.
B(5'): Mary also left.
A(6'): Hmm. So, do you know why?

b. why: why Mary left recently. or: why Mary left recently and Bill left recently. **not:** why does everybody hate Bill, why Bill left recently.

(21) contains similar data: after accepting B's initial assertion in (2), A's elliptical query in (3) will be resolved as (21a). However, if instead of (3), the dialogue continues as (3'-5'), the elliptical query in (5') cannot get the resolution in (21a), it can only get the resolutions listed in (21b):

(21) A(1): I've been gone a while. What's the news?
B(2): Millie was attacked.
A(3): hmm. Where?/When?

a. Where/when was Millie attacked
A(3'): mmh.
B(4'): Brendan was shot at.
A(5'): Oh gosh. Where?/When?

b. where/when was Brendan shot at; where/when was Millie attacked and Brendan shot at. **not**: Where/when was Millie attacked.

The obvious conclusion from these data is that the presupposition–motivated structure on FACTS is not sufficient to capture the resolution possibilities for fact-operator ellipsis.[10] As (20) illustrates, what is required cannot be global (certain SOA's that continue to be taken for granted, do not continue to serve as candidate fact–ellipsis resolvers.) In practice, for examples like (20) and (21) the commonest intended resolution is the fact provided by the *most recently accepted assertion.* This merely reinforces the anti-globalist point made here. Consequently, we *cannot*, as things stand, analogize the short-answer rule along the following lines:

(22) a. S $\rightarrow$ AdvP

b. Content(S)(DGB) = [CONTENT(AdvP)](f), where f $\in$ DGB|FACTS

As Asher 1993 emphasizes for abstract entity anaphora in texts, standard DRT and dynamic approaches are helpless when faced with such data—their most straightforward extension to these cases would predict that once a fact antecedent gets introduced into the context, it should, more or less, be available for ever more, in contradistinction to the data presented above.

4.2 The Right Frontier Constraint and Dialogue

Although the dialogue phenomena described in the previous section are, to the best of my knowledge novel, there has been quite a bit of work on anaphoric reference to propositional/fact entities in texts (see e.g. Polanyi 1987, Webber 1991, Asher 1993). Both Webber 1991 and Asher 1993 have proposed accounts for such data based on the right frontier constraint (RFC) proposed by Polanyi 1987, who originally proposed a parsing algorithm for discourses that incorporates the RFC.

I will assume as a working hypothesis that something like the RFC can be exploited to provide an account of fact ellipsis and anaphora potential in dialogue. However, there are two crucial differences between text and dialogue that lead me to consider a different formulation. For text just about any new unit needs to get attached to the discourse structure and constitute a fact antecedent, but at the same time the 'topic'–based structure is in many cases rather *abstract*—hence Polanyi and Asher need to "manufacture" topics (from individuals "about" which a text concerns). In dialogue, by contrast, a significant number of contributions don't constitute information, but rather are questions that get posed (or arise) and actually provide a fairly concrete notion of 'topic'. Fact antecedents emerge from information that accumulates about previously introduced questions—after acceptance (or accommodation).[11] The analogy I draw, nonetheless, relates the text-

[10] For additional data leading to this conclusion see Ginzburg 1997b, including data concerning "hasty accommodation", where one DP rashly adds a SOA into FACTS without waiting for acceptance from the other DP and then has to retract.

[11] An additional difference from text is that in dialogue information about the utterance needs to be kept track of for a minimal period, during grounding, e.g. for use in "metalinguistic" ellipsis, as we shall see in section 5.

derived notion of *open constituent* ("unexhausted topic") with the dialogue–derived notion of *question currently under discussion.*

Thus, the right frontier constraint translates informally into dialogue terms as follows:

(23) Antecedents for fact ellipsis and anaphora are those SOA's pertaining to questions currently under discussion.

To make this work, however, we need to revise our setup, to provide a notion of 'SOA pertaining to currently discussed question'.

4.3 A Revised Setup

The phenomena discussed in the previous section lead me to the following conclusion: one needs to recognize two grades of SOA's within DGB | FACTS. The first grade corresponds to the structure posited above: let us call such SOA's STORED. Such SOA's are to be thought of as items of information that truly have the acceptance of all conversational participants, following perhaps some discussion. They can thus be safely integrated with the conversationally emergent body of knowledge: FACTS | STORED will be closed under $\vee$ and $\wedge$.

The second grade in FACTS, to be dubbed TOPICAL, concerns SOA's that pertain to questions under discussion at that point in time. FACTS | TOPICAL will be updated *defeasibly* in such a way that *later accepted material takes precedence.* The hypothesis I make is that:

(24) It is precisely the SOA's in FACTS | TOPICAL to whom access by ellipsis and pronominal anaphora is possible.

More concretely, the revised set-up involves the following:

- FACTS contains SOA's of two sorts: TOPICAL and STORED.
- STORED: closed under $\vee$ and $\wedge$.
- TOPICAL: this will be treated as a set of pairs of $\langle$ question$_0$,soa$\rangle$, where question$_0$ is an element of QUD. Each such pair is such that the SOA stands in the ABOUT relation to the question. The question component of the ordered pair will be known as the ADDRESS: if $a = \langle q, \sigma \rangle$, then $address(a) = q$. Thus, the questions from QUD structure TOPICAL according to "subject matter". TOPICAL is updated using *priority union* (Grover et al. 1994), a defeasible update operation in which later accepted material takes precedence.[12]

[12] The motivation for this is to allow "facts" hastily accommodated into TOPICAL to be defeated by subsequent information before they get permanently into STORED. See Ginzburg 1997b for details.

The basic connection between TOPICAL and STORED is given by the revised version of the QUD Downdating Principle: this says that once a DP accepts into TOPICAL a goal fulfilling SOA, she can downdate the QUD–maximal question *and* "change the status" of the SOA's pertaining to that question in TOPICAL, making them "permanent", i.e. elements of FACTS | STORED:

(25) **QUD DOWNDATING PRINCIPLE [revised version]:** Assume q is currently maximal in A | QUD, and that there exists
$\langle q, \sigma \rangle \in$ A | FACTS | TOPICAL such that
$\sigma \Rightarrow_{UNPUB-MS(A)} goal - SOA(UNPUB - MS(A))$
Then, and only then permit DP to:

1. remove q from QUD:
   A | QUD:= A | QUD $\setminus\{q\}$
   $\prec := \prec \setminus \{\langle q_i, q_j \rangle \mid$ either $q_i =$ q or $q_j =$ q $\}$
2. Update FACTS | STORED:
   FACTS | STORED:= FACTS | STORED $+_{fact-closure}\ \sigma$

Unfortunately for reasons of space I cannot detail how querying and assertion work on this revised picture,[13] the basic idea relevant for current concerns is simply this: whereas in the view described previously, querying affected only QUD, here it will also affect FACTS: posing a question introduces a new element into QUD. In addition, it also introduces a new ADDRESS in TOPICAL about which SOA's can be provided. Technically, this is implemented by introducing into TOPICAL a pair consisting of q and the vacuous SOA $\top$. At the same time, I also assume that when a new question gets introduced, the addresses for questions that are no longer under discussion are downdated from TOPICAL. This latter assumption, following (23), represents our own version of the right frontier constraint:

(26) a. q becomes maximal in A | QUD.

b. A | FACTS | TOPICAL:= FACTS | TOPICAL $\cup \langle q, \top \rangle$
$\setminus \{a \mid address(a) = q \wedge q \notin QUD\}$

4.4 Fact ellipsis

To get some idea how this revised set-up works let us consider how data like that in (20) can be accounted for. The one additional step we need to make is to postulate an interpretive schema for such uses. Given our assumption in (24), the following schema is quite natural:

(27) a. S $\rightarrow$ AdvP

b. Content(S)(DGB) = CONTENT(AdvP)(f),

c. Context: $\exists \alpha \in$ DGB | FACTS | TOPICAL and $\alpha = \langle q, f \rangle$, for some q.

[13] For which see Ginzburg 1997b .

The move by move analysis in (28) abstracts away from various details irrelevant for current concerns, concentrating solely on the evolution of TOPICAL. I use notation like 'Bill?' as a transparent (in context) name for a question, here the question expressed by 'whether Bill has left recently':

(28) A: Who's left recently?
B: Bill.
A: Aha. Why?
(1) A: Who's left recently?
FACTS | TOPICAL: $\langle q_1, \top \rangle$

(2) B: Bill.
FACTS | TOPICAL: $\langle q_1, \top \rangle$, $\langle Bill?, \top \rangle$

(3) A: Uh huh.
FACTS | TOPICAL: $\langle q_1, Bill \rangle$, $\langle Bill?, Bill \rangle$

Downdates Bill?, q_1? from QUD: QUD:= $\emptyset$
(4) A: Why?
FACTS | TOPICAL:= $\langle q_2, \top \rangle$

The dialogue works essentially as follows: A asks a question in (1) to which B responds in (2). A accepts the assertion in (3). Let us assume she is now ready to move on to another issue, the one she raises in (4), so she can now downdate from QUD both her original question and the question designated in (28) as 'Bill?' At this point there is one possible fact antecedent in TOPICAL, 'Bill left recently'. This can, therefore, serve as an antecedent for A's 'why' in (4). However, a side effect of A's posing her question in (4) is that the addresses corresponding to Bill?' and to the A's initial question get downdated from TOPICAL, since these questions are no longer in QUD. Notice that there is, thus, always a one move lag between the downdating of questions from QUD and the disappearance of the addresses they provide in TOPICAL. This seems like an intuitive prediction: once some information is no longer contentious, one still wants to be able to use it as a constituent of other contents which "comment" on it.

5 Interacting over Utterances

Let me now add the final component to the picture: the framework sketched sofar has allowed for illocutionary actions (querying and assertion) but has ignored completely the fact that these acts typically involve the occurrence of an utterance. Of course such inattention is the norm in the lion's share of semantic practice, formal or otherwise. While this is understandable as far as text goes, where the communicative process is by and large an invisible one,[14] for dialogue this is unacceptable since a large proportion of dialogue activity actually relates to the conversation itself—whether the previous utterances have

[14] Though some writers attempt to make it visible, the paradigm case being perhaps Wittgenstein.

been understood; if not, what needs clarifying and so forth.

How does the fact that NL utterances involve a communicative process affect context? The short answer I will give here takes as its starting point an account of dialogue interaction due to Clark and his collaborators. (e.g. Clark and Schaefer 1993). Their view can be summarized very roughly in terms of the following generalization:

(29) *(After each contribution in dialogue) The contributor and his partners attempt to satisfy the grounding criterion: the contributor and his partners mutually believe that the partners understood what the contributor meant relative to their own purposes.* (Clark and Schaefer 1993 p. 148.)

This process is illustrated in (30):

(30) A: Uhm...Now on this map , if you were to move upwards slightly and to the right and then once you'd moved to the right til about three quarters along the paper.

B: Mhm.

A: And then move upwards.

B: Uhuh.

A: You would come to a ruined monastry.

B: Uhuh.

A: Is that the same.

B: Yeah.

(From the Map-Task)

Allowing for a communicative process involves in the first instance recognizing that *the content of an utterance by A is not automatically transparent to the participants of the dialogue apart from A.* The initial move I make in order to effect such recognition is to reify an utterance as an event/situation spatio-temporally located in the world, a move initially motivated in Barwise and Perry 1983 and developed in subsequent situation semantics and HPSG work. The main consequence of this reification is that an utterance, like any other spatio-temporally located entity, can be the object of description or, alternatively, wondering: agents can come to agree on how to describe the utterance and can pose questions concerning it. A second component of my account concerns QUD: I suggest that one way of operationalizing the grounding strategy of (29) is to assume that A making an utterance u raises an issue for discussion *for B*, which I shall refer to as u's grounding question *g-q(u,B)?*. The issue can be formulated, very much in the spirit of Grice as follows:

(31) whether B, the addressee, understands what A, the utterer, meant by the utterance u.

The consequence of this is that the first issue B has to attend to on forming the belief that an utterance has been made is u's grounding issue. Broadly speaking, the addressee will have the option of explicitly acknowledging comprehension ("accepting the utterance") which will commit her to the belief that the grounding criterion has been met; this will also involve the content of the utterance getting added into the facts component of the common ground; alternatively, she will have the option of starting a discussion of (31) or related issues.

The interaction protocol for grounding based on (31) will involve a basic asymmetry between the utterer and the addressee: as far as the addressee goes, I assume that after any utterance u, more precisely—as soon as she forms the belief that an utterance was made,[15] she must always initially consider g-q(u,B)?. There is no choice in the matter; there is a distinct contrast here with the case of queries/assertions, where an addressee can coherently fail to accept a question for discussion:

(32) a. I don't know./Why do you think I know?

b. Do we need to talk about this now?/ O.K. let's talk about this now.

c. I don't wish to discuss this./Why should I talk about this with you?

Once an utterance is made, however, the addressee *must* attend to its having taken place. As for the speaker—I will assume that as a rule g-q(u,B)? is not considered by A, that is—it does not get added to A's QUD. This assumption is of course far from obvious: Ginzburg 1995b made the opposite assumption—primarily because of the apparent analogy with assertion.[16] However, as I have pointed out already for the addresee, there is actually a disanalogy with assertion: modulo hallucinations A *knows* what she has just said and meant and can assume that B will (eventually ...) know it too. Thus, it seems a more plausible strategy to assume that in general A herself has no explicit intention to discuss g-q(u,B)?—as far as her own DGB goes she ignores this issue.

5.1 Relativized Utterance Understanding

Now as (31) stands, it remains somewhat mysterious insofar as no explication of the notion 'B understands what A meant by u' is provided. In Ginzburg 1997c I argue that the understanding task facing an addressee can be modelled in terms of questions she needs to find an answer for, the *content question* and the *goals question.* Specifically:

(33) a. A makes an utterance u of S whose meaning is μ.

b. The addressee B faces two questions:

1. $(u?\mu)$: 'what values does u provide for the variables of μ'
2. GOALS(A,u)?: 'what goals does A intend to achieve **by** making the utterance u'.

c. **Resolvedness as a criterion for grounding:** given an agent B with mental situation UNPUB-MS(B), B grounds u iff B believes it is mutually believed that he found facts τ, σ such that:
τ resolves $(u?\mu)$ relative to UNPUB-MS(B) AND
σ resolves GOALS(A,u)? relative to UNPUB-MS(B)

[15] And more precisely and realistically still: during the event itself. I will not, however, be considering incremental interpretation here.

[16] Indeed there might be some motivation for adopting this latter position in a more incremental version of the current framework that was concerned also with speaker repair. More plausibly though, the speaker considers a somewhat different question about her own performance but this is a matter that I will not consider here.

5.2 Uttering an Assertion

How to operationalize the interaction on grounding? I will restrict myself here to a very simplified account of the utterance interaction associated with asserting, just enough to explain the source of some basic mismatches as to the availability of semantic objects; the case of querying is quite analogous.

(34) Uttering an assertion

- A makes u:
  - Updates QUD with p?
  - A | FACTS | TOPICAL:= FACTS | TOPICAL $\cup \langle p?, \top\rangle$ $\setminus \{a \mid address(a) = q \wedge q \notin QUD\}$
  - A | LATEST-MOVE: τ—a fact that resolves $(u?\mu)$, the content question associated with u.
- B realizes an utterance was made:
  - g-q(u,B) is maximal in B | QUD.
    FACTS | TOPICAL:= FACTS | TOPICAL + $\langle g - q(u, B), \top\rangle$ $\setminus \{a | address(a) = q \mid q \in QUD\}$
- Option 1: **Grounding:**
  1. **B makes an *affirmative* utterance u_1 about g-q(u,B)**
  2. **B increments her FACTS:**
     FACTS — LATEST-MOVE := σ—a fact that resolves $(u?\mu)$.
     FACTS | TOPICAL:=
     FACTS | TOPICAL + $\langle$ $(u?\mu)$,$\sigma\rangle$, where σ is as above.
  3. **B downdates g-q(u,B) from QUD**:
  4. **B reacts to the assertion: accepts it or discusses it.** (See (13) above.)
- Option 2: **Clarify g-q(u,B)**: B provides a response that addresses g-q(u,B):
  - If content-question(u) is unresolved, B can add as maximal in QUD any sub-question of content-question(u).
  - If Goals(u) is unresolved, B can add as maximal in QUD any sub-question of Goals(u).

(34) represents the explicitly asymmetric position I have suggested utterer and addressee stand in. When compared with the previous view of assertion, given in (13), which abstracted away from communicative interaction, we see that there is no essential change as far as the utterer goes. For the addressee, however, a whole new layer of action has been postulated: this starts with the necessity of updating QUD with g-q(u,B); this, in turn, has the consequence that facts that resolve the content question, $(u?\mu)$, associated with g-q(u,B)? have the potential of being TOPICAL. It also begets the potential either for affirmative utterances that commit the addressee to being able to ground the utterance (as in (30)) or to posing clarification questions about the utterance.

6 Revisiting the Turn-Taking Puzzle

Let me now apply the combined force of the framework sketched above to tackle the TURN TAKING PUZZLE. Consider once more (2) repeated as (35):

(35) a. A: You're upset. Why? (unambiguously: 'Why is B upset')

b. A: You're upset. B: Why? (Strong preference for: 'Why is A saying that B is upset'; weak possibility of: 'Why is B upset')

As for (35a): the reading for 'why' that A can get is possible since, given the setup of section 4, A can accommodate a positive resolution of 'whether B is upset' into her FACTS | TOPICAL after making her assertion; this gives her the requisite fact. On the other hand, since g-q(u,B)? never gets into A's QUD, a fact positively resolving ($u?\mu$)) is not topical, so cannot be manipulated.

The situation with (35b) is almost the converse: given that g-q(u,B)? is initially in B's QUD, a fact positively resolving ($u?\mu$) is topical, so can be manipulated, yielding the preferred reading of (35b). The response as a whole is coherent since it is a clarification of the u's GOALS question. After grounding has taken place a positive resolution of 'whether B is upset' becomes topical, and hence available for resolution.

Now for the other version of the puzzle:

(36) a. A: Who solved the chess problem? (and) How? (= How did a solver solve the chess problem?)

b. A: Who solved the chess problem? B: How? (= What means of solution$_m$ are you asking who solved the problem by doing m?)

The explanation I offer here is quite similar. As far as (36a) goes: A can decide to accommodate a positive resolution of *who solved the chess problem* into her FACTS | TOPICAL; this gives her the requisite fact. On the other hand, since g-q(u,B) never gets into A's QUD, a fact positively resolving ($u?\mu$)) is not topical, so cannot be exploited as a null anaphor.

As for (36b): for B a fact positively resolving ($u?\mu$)) *is* topical, so can be used in resolution. Why is the (36a) resolution not possible? *Before* B grounds A's question, the posed question q it is not in B's QUD. Hence, accommodating a fact positively resolving q, pos-res(q), is not possible. *After* grounding there are a couple of options:

- If an answer is provided, this answer will subsume the fact that is pos-res(q). So the pos-res(q) reading disappears.
- If no answer is provided, the pos-res(q) reading remains as an option—since q does not automatically disappear from B's QUD if B fails to accept q.[17] This is illustrated in (37):

[17] For instance: A: Who left?/B: I'm not sure./A: Someone we know?/B: Maybe Jill.

(37) a. A: Who solved the chess problem?

b. B: Gary and Judit. I know how too.

c. B: I'm not quite sure. I do know how though.

7 Conclusions

In this paper I have sketched an approach to dialogue where

- Illocutionary acts and utterance acts can be treated in a uniform manner.
- Where facts receive both a long-term structuring, as well as a temporary one determined by the questions currently under discussion.

I have motivated this approach by considering certain data concerning factive operators, which can exploit facts that arise in context from either type of act. Moreover, I have argued that any model of dialogue needs to accommodate the following:

- In dialogue there is no equal access to the semantic objects.
- A semantics for dialogue needs to pay attention to the different roles of DP's at different points in a conversation.
- Certain structure, particularly that relating to utterance acts, is "imposed" on the DP's.

References

Asher, N.: 1993, *Reference to abstract objects in English: a Philosophical Semantics for Natural Language Metaphysics*, Studies in Linguistics and Philosophy, Kluwer, Dordrecht

Barwise, J. and Perry, J.: 1983, *Situations and Attitudes*, Bradford Books, MIT Press, Cambridge

Clark, H. and Schaefer, E.: 1993, 'Contributing to Discourse', in *Arenas of Language Use*, pp 259–94, CSLI Publications, Stanford

Cooper, R. and Ginzburg, J.: 1996, 'A Compositional Situation Semantics for Attitude Reports', in J. Seligman and D. W. hl (eds.), *Logic, Language, and Computation*, CSLI Publications, Stanford

Crimmins, M.: 1993, *Talk about Beliefs*, Bradford Books, MIT Press, Cambridge

Ginzburg, J.: 1995a, 'Resolving Questions, I', *Linguistics and Philosophy* 18, 459–527

Ginzburg, J.: 1995b, 'The Semantics of Interrogatives', in S. Lappin (ed.), *Handbook of Contemporary Semantic Theory* , Blackwell: Oxford

Ginzburg, J.: 1996, 'Semantic ellipsis resolution with syntactic presuppositions', in H. Bunt (ed.), *Proceedings of the 2nd International Workshop on Computational Semantics*, ITK:Tilburg

Ginzburg, J.: 1997a, 'Querying and Assertion in Dialogue, I', *Hebrew University of Jerusalem Ms.*

Ginzburg, J.: 1997b, 'Querying and Assertion in Dialogue, II: defeasible presuppositions, fact ellipsis, and non-cooperativity', *Hebrew University of Jerusalem Ms.*

Ginzburg, J.: 1997c, 'Understanding in Dialogue', *Hebrew University of Jerusalem Ms.*

Grover, C., Brew, C., Manandhar, S., Moens, M., and Schoter, A.: 1994, 'Priority Union and Generalization in Discourse Grammar', in A. Schöter and C. Vogel (eds.), *Edinburgh Working Papers in Cognitive Science: Nonclassical feature system* , Centre for Cognitive Science, University of Edinburgh

Karttunen, L.: 1977, 'Syntax and Semantics of Questions', *Linguistics and Philosophy* 1, p. 3–44

Lewis, D. K.: 1979, 'Score Keeping in a Language Game', in R. Bauerle (ed.), *Semantics from different points of view*, pp 172–187, Springer

Polanyi, L.: 1987, 'A formal model of the structure of discourse', *Journal of Pragmatics* 12, 601–638

Roberts, C.: 1995, 'Information Structure in Discourse', *Ohio State University Ms.*

Stalnaker, R. C.: 1978, 'Assertion', in P. Cole (ed.), *Syntax and Semantics, Volume 9*, pp p. 315–332, AP, New York

Webber, B.: 1991, 'Structure and Ostension in the Interpretation of Discourse Deixis', *Language and Cognitive Processes* 6, 107–135

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

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)
$$\begin{array}{ccc} \langle\langle ON, cat, mat; 1\rangle\rangle & \Longrightarrow & \begin{bmatrix} \text{phon: } \textit{the cat is on the mat} \\ \text{synsem: } \begin{bmatrix} \text{content: p} \end{bmatrix} \end{bmatrix} \\ \Vert & & \\ s^1 & & \end{array}$$

[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.

$$\text{(4)} \qquad \underset{\substack{\Downarrow \\ s^1}}{\langle\langle ON, cat, mat; 1\rangle\rangle} \implies \underset{\substack{\Downarrow \\ s^2}}{\begin{bmatrix} \text{phon: } \textit{the cat is on the mat} \\ \text{synsem: } \begin{bmatrix} \text{content: p} \end{bmatrix} \end{bmatrix}}$$

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

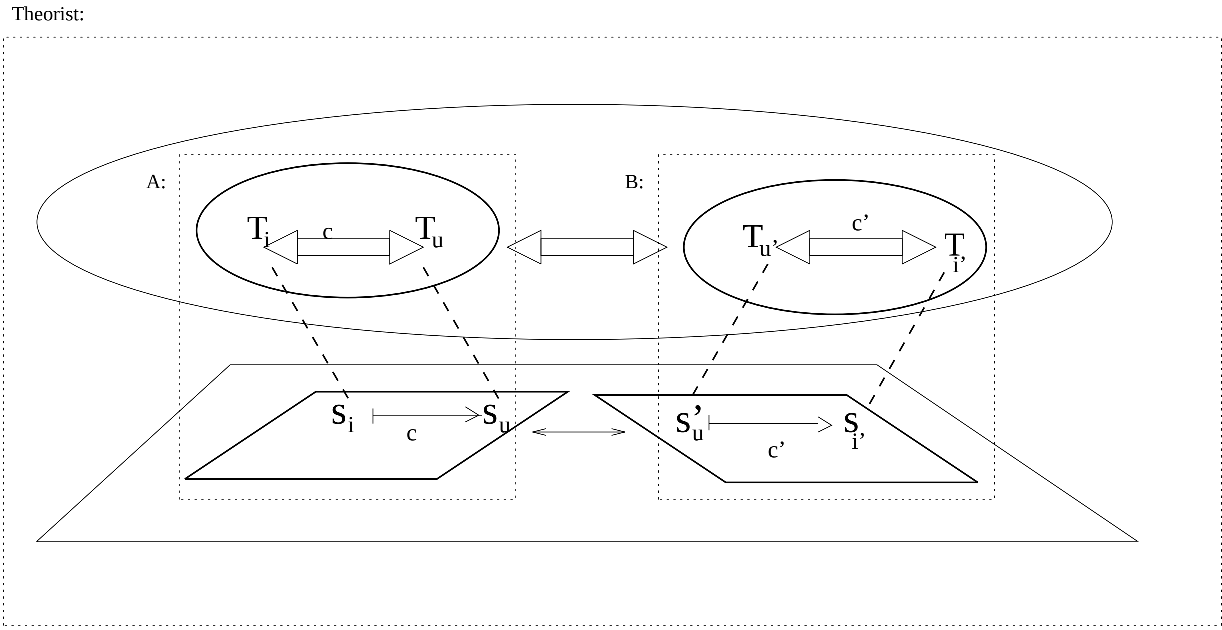


Fig. 1: Inter-Agent Signal

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

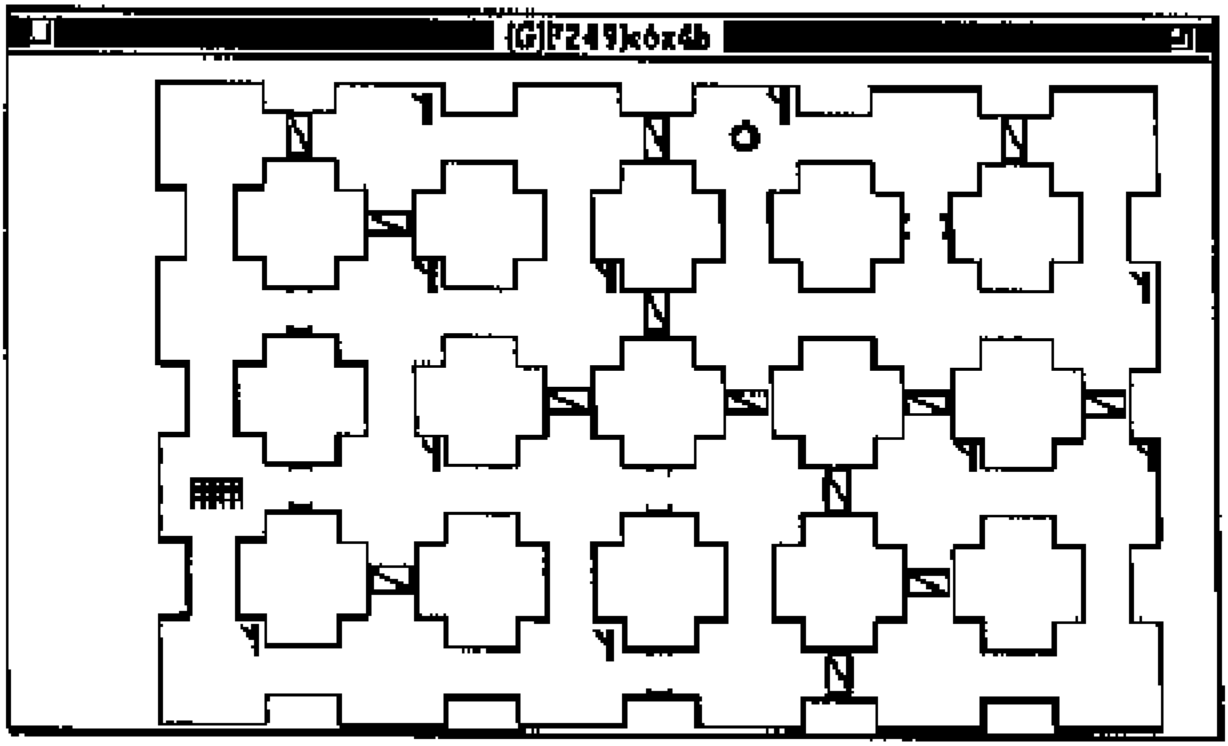


Fig. 2: The circle marks the location P_A is talking about.

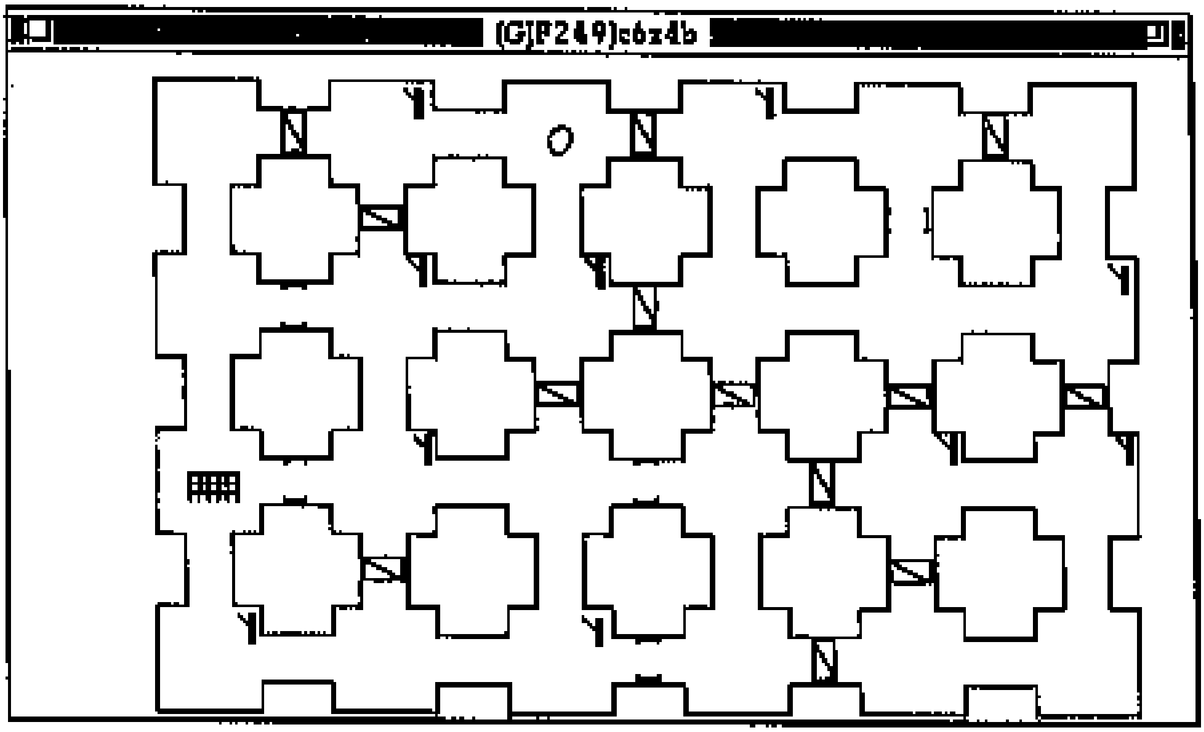


Fig. 3: The circle marks the location P_B understands P_A to have indicated.

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)
$$
\begin{array}{ccc}
\begin{array}{|ccc|}
\hline
I & \Longleftrightarrow & U \\
\rotatebox{90}{$\models$} & & \rotatebox{90}{$\models$} \\
s^1 & \overset{c'}{\longleftrightarrow} & s^2 \\
\hline
\end{array}
&
\begin{array}{||c||}
\hline
\Longleftrightarrow \\
\\
\overset{c''}{\longleftrightarrow} \\
\hline
\end{array}
&
\begin{array}{|ccc|}
\hline
U' & \Longleftrightarrow & \neg I \\
\rotatebox{90}{$\models$} & & \rotatebox{90}{$\models$} \\
s^3 & \overset{c'''}{\longleftrightarrow} & s^4 \\
\hline
\end{array}
\\
\mathrm{P}_A & & \mathrm{P}_B
\end{array}
$$

[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)
$$
\underbrace{\begin{array}{|ccc||c||ccc|}
\hline
I & \Longleftrightarrow & U & \Longleftrightarrow & U' & \Longleftrightarrow & \neg I \\
\rotatebox{90}{$\models$} & & \rotatebox{90}{$\models$} & & \rotatebox{90}{$\models$} & & \\
s^1 & \xleftrightarrow{c'} & s^2 & \xleftrightarrow{c''} & s^3 & & \\
\hline
\end{array}}
\qquad \mathrm{P}_A \qquad\qquad \mathrm{P}_B
$$

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)
$$
\underbrace{\begin{array}{|ccc||c||ccc|}
\hline
I & \Longleftrightarrow & U & \Longleftrightarrow & U' & \Longleftrightarrow & I \\
\rotatebox{90}{$\models$} & & \rotatebox{90}{$\models$} & & \rotatebox{90}{$\models$} & & \\
s^1 & \xleftrightarrow{c'} & s^2 & \xleftrightarrow{c''} & s^3 & & \\
\hline
\end{array}}
\qquad \mathrm{P}_A \qquad\qquad \mathrm{P}_B
$$

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 Co-ordination 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.

Structured Information States*
– raising and resolving issues –

Joris Hulstijn†

Abstract

This paper proposes to combine update semantics with a partition semantics of questions. The resulting *update semantics with questions* (USQ) models a question as a change to the information-structure of an information state. Each information state is considered to be structured by one or more *issues*: salient questions in need of an answer. This makes it possible to clarify the notion of *relevance*. An utterance U is considered relevant when it (partially) resolves one or more of the current issues. For automatic inquiry systems some ways of predicting the currently salient issue, are indicated.

1 Introduction

Every day we are bombarded with a multiplicity of information: books, magazines, conversation, fax and the internet might result in an 'information overkill'. Yet we are able to structure such large amounts of data. We select precisely the information that meets our needs, answers our questions or helps to solve our problems. In short, information that is *relevant*. For discourse or dialogue such a picture suggests the following *problem of relevance*:

> When can an utterance U with information content ϕ by called *relevant* in discourse?

This paper attempts a formal characterization of a notion of relevance in cooperative discourse (Grice 1975). The aim is to develop a theory of information exchange for goal-directed discourse. It should at least express the difference between relevant and non-relevant utterances. The key idea is to think of information as data that is somehow structured.

information = data + structure

* Thanks to Jeroen Groenendijk, Henk Zeevat, Jelle Gebrandy and Gerhard Jäger for comments and corrections. All errors remain mine.

† Computer Science, University of Twente, PO BOX 217, 7500 AE Enschede, The Netherlands, (31)(53)4894652, joris@cs.utwente.nl

Following Ginzburg (1995) and Van Kuppevelt (1995) I assume that information in a dialogue-context is structured by one or more *questions under discussion* or *issues*: salient questions that have not yet been answered at that point in discourse. Suppose that all questions that need to be answered, or problems that need to be solved can be represented in terms of such issues. Then we might say that information that does not contribute to *resolving* the issues, can not be called relevant.[1]

> An utterance U can be considered *relevant* when it's information content ϕ fully or partially resolves one of the current issues.

Given this basic idea, a number of notions needs to be specified. How do we model the information content of utterances? How do we model the structure of information? What is an issue? When can we say that information in an utterance completely or partially resolves an issue? How do we find out what the current issues are, at a given point in discourse? Can we predict what the issues will be, from the way the discourse develops?

1.1 Update Semantics with Questions

I will use an *update semantics with questions* (USQ) to specify the notions of information and issue and to provide an account of what it means to – fully or partially – resolve an issue. USQ is a combination of update semantics (Veltman 1996) and a partition semantics of questions (Groenendijk and Stokhof 1984; Groenendijk and Stokhof 1996). A similar combination is proposed by Jäger (1996) and Groenendijk (1997).

Update semantics models the increase of information as the elimination of epistemic possibilities in an information state (cf. Stalnaker 1979). A partition semantics of questions models a question as a partition or equivalence relation over the epistemic possibilities in an information state. Each equivalence class corresponds to one of the exhaustive answers to the question. The combination of the two, USQ, models a question as a change to the information structure of an information state. As suggested above, each information state is considered to be structured by one or more *issues*[2]. Like a 'grid', an issue may structure information. It specifies classes of epistemic possibilities. Possibilities in one class may differ on irrelevant facts, but share the relevant facts. Information is said to partially resolve an issue when it eliminates all epistemic possibilities from at least one of the alternative classes.

This view on information with two dimensions is not a new or alien view. In structured information states, information can be increased in two ways: by eliminating epistemic possibilities, thereby increasing data, or by extending or refining the structure of the issues. This aspect is reminiscent of work by C.S Peirce (1867). A view of information as a

[1] The notion of information *resolving* an issue is an extension of the notion of an assertive utterance answering a question. Resolution depends on the information state and on the goal of the user. See (Ginzburg 1995).

[2] Groenendijk and Stokhof use the term *question* for the semantic object that corresponds to an *interrogative*. I use 'question' in its colloquial sense. My technical term is *issue*. Issues can be raised in other ways than by asking questions explicitly.

set of relevant alternatives is actually quite close to ideas expressed by Stalnaker in his book *Inquiry* (1984). Mental representation enables reasoning about actions and goals. Deliberating outcomes of possible actions makes it possible to choose the best possible alternative. Such ideas can be found in research from AI. Decisions are made by estimating the benefits and costs of alternatives and by comparing the importance of conflicting goals. See also section 3.3.

1.2 Issues and Goals

It is notoriously hard to derive or predict the current issue or issues at a given point in discourse.[3] Only after the fact, one can work out the implicit question that actually prompted an utterance. For instance in example (1), adapted from (Kuppevelt 1995), general constraints about news stories or terrorist actions suggest follow-up issues. Also, new issues may be raised by general narrative constraints. For instance about the consequences of an event. However, the order in which possible issues are addressed and whether they are important enough to be addressed at all differs for each case.

(1) Yesterday evening a bomb exploded outside the Houses of Parliament. (*victims?*) Nobody was killed. (*who did it?*) The attack was claimed by separatist rebels of the Liberation Army. (*consequences?*) The prime minister condemned the assault, adding that further attacks would endanger the peace process.

Although it is not possible to predict issues in general, for a fixed task and a fixed domain heuristics can be given to predict the most likely issue. One such task domain is dealt with in the SCHISMA system. SCHISMA is an automatic inquiry system for theatre information (Hoeven et al. 1995). By the very set-up of the system we may assume that the user will have external perlocutionary *goals*: to go to the theatre, get information about performances and possibly order tickets. Such goals may help to predict issues.[4] For automatic inquiry systems like SCHISMA I will discuss a number of ways in which issues can be raised.

Assuming an issue can be predicted, USQ may be of use in theoretical linguistics analyzing the following puzzling phenomena: domain restriction, semantics of focus and nominal anaphora resolution. All of these have to do with relevance in some way or other. I would like to stress that issues are not only of importance to interrogatives, but also to seemingly simple declaratives.

The remainder of the paper consists of two parts. The first part will be rather technical. It contains the definitions of USQ and specifies the notions of information state, issue and what it means to resolve an issue. The second part deals with determining and predicting the current issue. Although this problem is not solvable in general, for a concrete case

[3] There can be more issues at work at the same time. Ginzburg postulates a partial order among issues, indicating a *salience*. I have no opinion on salience orderings, apart from the obvious information order (section 2.3). For inquiry systems issues stand in a hierarchical order following some pre-defined topic structure (section 3.2).

[4] I agree with Ginzburg (1995) that user goals are necessary in any theory of answerhood.

such as the SCHISMA system it can indeed be solved. By reducing relevance to issues, and issues to domain-driven goals, you might say that the 'linguistic' problem of relevance, can be reduced to an 'engineering' problem: the problem of modelling a given task domain as precise as possible. The paper ends with a summary and discussion of future research.

2 Update Semantics with Questions

Systems of update semantics have successfully dealt with phenomena on the border of semantics and pragmatics. One example is presupposition theory. (Beaver 1996) Presuppositions come back later this paper. Another example is the treatment of anaphora and indexicals. (Groenendijk et al. 1996) Compared to that I simplify things a bit. In this section I present a version without variables and quantification. In section 2.4 quantification will be added. For a proper account of the relation between anaphora and questions see (Groenendijk 1997).

Update Semantics is a logical theory that specifies how an information state changes during discourse. The meaning of an utterance is seen as a transition between information states. In this paper, information states are not used to model the beliefs, commitments or knowledge of a single participant in dialogue, but rather the discourse context according to an objective observer. The observer does not take part in the conversation, but monitors the information as it is publicly available from the utterances, or can be assumed to be common background. The reason for this is that for questions at least two participant roles have to be considered. The asker of the question, who does not know the answer, is obviously in a different information state than the answerer of the question. For more on the interesting question how the public or common discourse context relates to the beliefs or commitments of individual participants I refer to (Zeevat 1997).

Unfortunately, part of the 'sting' of relevance, relevance for somebody, is taken out in this way. In a later stage USQ information states can be relativized to asker and answerer roles. The important research topics of misunderstanding and corrections will also have to be postponed.

As in most epistemic logics, it is assumed that information is to remain consistent at all time and stay closed under logical consequence. For dialogue participants such rationality assumptions are known to be wrong. The context may get inconsistent or participants may not be able to grasp all the consequences of what has been said. For a psychologically plausible account of discourse context I refer to (Walker 1996).[5]

The basic idea of update semantics is that increasing information means narrowing down a set of epistemic possibilities. Those possibilities that are inconsistent with newly acquired information are eliminated. Think of epistemic possibilities as possible ways the world can be, according to given information. Epistemic possibilities can be thought of as

[5] Walker successfully attacks the rationality assumptions. In particular she attacks the *redundancy constraint* that utterances containing information already present in the discourse context are redundant. Empirical evidence shows that people often use informationally redundant utterances, for instance to acknowledge acceptance or signal the use of an inference rule.

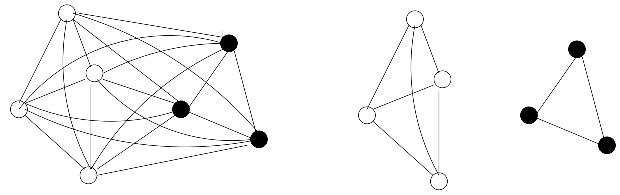


Fig. 1: $u(A)$ and the partition $A/same\_colour$

complete possible worlds, but they certainly need not be. For a dice game, there would be 6 interesting possible ways the world can be. It is possible to compare epistemic possibilities to situations (section 4).

2.1 Definitions

Formulas of the logical language represent the information carried by an utterance. The utterance need not be a sentence, but could be a single word or a gesture. I assume that a grammar translates utterances to meaningful formulas. A formula is interpreted as a transition between information states. Formally an Update Semantics for a logical language L is a *frame* $\langle \Sigma, \{[\phi]\}_{(\phi \in L)} \rangle$, consisting of a set of information states Σ and a partial update function $[.]$ over Σ.[6] Following Veltman (1996) a postfix notation is used: $\sigma' = \sigma[\phi]$ means that σ' is the result of updating an information state σ with a formula ϕ.

At first the logical language L is taken to be the language of predicate logic without quantification, extended with a question operator ? and with a presupposition operator ∂. Individual constants are used as *referents*. Quantification will be added later. (section 2.4) The set of ground formulas is of central importance in the definitions. It is called the *vocabulary* $\mathcal{A}$.

Definition 1 (Language (L)): Given sets of n-ary predicates $Pred^n$ and referents $Refs$, define

$$
\begin{array}{lll}
\mathcal{A} & = & \{Pa_1...a_n \mid P \in Pred^n, a_i \in \text{Refs} \;\; (1 \leq i \leq n)\} \\
L & ::= & p \mid \neg\phi \mid \phi \wedge \psi \mid \phi \vee \psi \mid \phi \rightarrow \psi \mid \phi \Rightarrow \psi \mid \partial\phi \mid ?\phi \;\; (p \in \mathcal{A}, \phi, \psi \in L)
\end{array}
$$

In normal update semantics information states are defined as *sets of epistemic possibilities*. Technically an epistemic possibility w is a total valuation with respect to the ground formulas $\mathcal{A}$. So the set of all possibilities W consists of all mappings of ground formulas to 1 or 0. W may be called the logical space. In USQ elements from W are said to represent the *data* dimension of information. The basic epistemic possibilities contain all the detailed basic facts.

In order to model the *structure* dimension of information, we need some notions from algebra. An equivalence relation R over set A generates equivalence classes $\mathbf{a}_R$. In a graph

[6] The partiality is the result of presuppositions. A failing presupposition corresponds to an undefined result of the update.

the equivalence classes are the 'totally connected' subgraphs. See figure 1. The non-empty equivalence classes form a partition, called the *quotient set* of A over R, A/R. Starting from the partition we can in turn induce an equivalence relation, by taking all pairs of elements that are in the same block of the partition. Therefore it does not matter whether an equivalence relation or a partition is used to model structure.[7] One can always be turned into the other. The universal relation on A, written $u(A)$, combines all elements into a partition of one single block.

Definition 2 (Algebra): Given set A, relation $R \subseteq A \times A$, define

$\mathbf{a}_R$	$=$	$\{b \in A \mid \langle a, b\rangle \in R\}$ $(a \in A)$	equivalence class
A/R	$=$	$\{\mathbf{a}_R \neq \emptyset \mid a \in A\}$	quotient set
$u(A)$	$=$	$A \times A$	universal relation

Figure 1 depicts the partition of a set of coloured nodes A, induced by the relation $same\_colour$. All connections between nodes that do not 'agree on colour' have been eliminated. Likewise in USQ information is viewed as the quotient set of W over the equivalence relation induced by the current issues. All possibilities that 'agree on the relevant facts' are in one equivalence class.

The following definitions are largely based on Jäger (1996). The *data* dimension of information, $d(\sigma)$, is modeled by the separate possibilities. $d(\sigma)$ can be seen as a normal US information state, containing *declarative* information. The *structure* dimension with the *issues*, $i(\sigma)$, is modeled by the partition. This captures the *interrogative* part of information. The information state itself is modeled by an equivalence relation over the logical space, $\sigma \subseteq W \times W$. From σ both $d(\sigma)$ and $i(\sigma)$ can be derived. Σ is the set of all information states. We assume an initial 'ignorant' state $\mathbb{1} = W \times W$. It contains all epistemic possibilities and is therefore not structured.[8] We also assume a unique inconsistent or *absurd* state, $\mathbb{0} = \emptyset$. It has no structure either.

Definition 3 (Dimensions):

$d(\sigma)$	$=$	$\{w \mid \langle w, w\rangle \in \sigma\}$	data, declarative
$i(\sigma)$	$=$	$d(\sigma)/\sigma$	issues, interrogative

The semantics of USQ is defined inductively on all $\phi \in L$. Each of the clauses in the definition is explained below.

Definition 4 (USQ): USQ is defined by $\langle \Sigma, [\phi]_{(\phi \in L)}\rangle$, where

$$
\begin{array}{lcl}
\sigma[p] & = & \sigma \cap u(\{w \in d(\sigma) \mid w(p) = 1\}) \quad (p \in \mathcal{A}) \\
\sigma[\neg\phi] & = & \sigma \cap u(\{w \in d(\sigma) \mid w \notin d(\sigma[\phi])\}) \\
\sigma[\phi \wedge \psi] & = & \sigma[\phi][\psi] \\
\sigma[\phi \Rightarrow \psi] & = & \{\langle v, w\rangle \in \sigma[?\phi] \mid \text{ if } \langle v, w\rangle \in \sigma[\phi], \text{ then } \langle v, w\rangle \in \sigma[\phi][\psi]\} \\
\sigma[?\phi] & = & \{\langle v, w\rangle \in \sigma \mid v \in d(\sigma[\phi]) \text{ iff } w \in d(\sigma[\phi])\} \\
\sigma[\partial\phi] & = & \sigma[\phi], \text{ if } \sigma[\phi] \neq \mathbb{0} \\
 & & undefined, \text{ otherwise}
\end{array}
$$

[7] Partitions are used by Groenendijk (1997). I prefer equivalence relations.

[8] In applications $\mathbb{1}$, like all information states, will be constrained by meaning postulates representing domain knowledge.

For atomic formulas p, all possibilities incompatible with p are eliminated. Negation is usually modeled by set complement. However, information states are supposed to be equivalence relations. An expression like $\sigma - \sigma[\phi]$ would remove for instance all reflexive instances from σ so that it no longer represents an equivalence relation. The use of $\cap$ in the definition makes sure that negation is a socalled declarative update (see later), like atomic formulas.

Conjunction is modeled by function composition on updates. This gives a sequential notion of conjunction. Because of presuppositions conjunction is no longer commutative.[9] Implication is defined by the standard equivalence $\phi \to \psi \equiv \neg(\phi \wedge \neg\psi)$. Because the clause for negation is declarative, i.e. no structure can be induced under the scope of a negation, we would not be able to model conditional questions. But conditional questions give an interesting structure, see example (2). Therefore an alternative definition for conditionals is suggested, $\Rightarrow$, in which the antecedent must become an issue[10]. Only for those possibilities where the antecedent holds, the consequent is dealt with. This definition will be used for domain restriction in quantificational structures later on.

The clause for presupposition is based on (Hulstijn 1996). I put it in to show how USQ can be combined with existing proposals for other discourse phenomena. Proper names, definites, factive verbs, some lexical expressions and sometimes other quantificational expressions like wh-questions are taken to be *presupposition triggers*. Use of such a trigger indicates that the presupposition that is conventionally associated with the trigger is supposed to be part of the context already. A definite like *the president* would be modeled by the formula $\partial president(a_5)$, where a_5 is a referent. Uniqueness is neglected. A presupposition $\partial\phi$ can be treated in one of three ways: (i) the presupposition is already supported by the context. (notation $\sigma \Vdash \phi$) In that case nothing happens. (ii) The presupposition can be added consistently to the context. In that case, it will indeed be added or *accommodated*. (iii) The presupposition is inconsistent with the context. In that case, the utterance was inappropriate in that context. This is modeled by the resulting update becoming *undefined*. Presuppositions are the only reason that the update function is a partial function.[11]

The following example shows the structure of questions depicted as a grid. The question 'whether p' is depicted by a vertical division, the question 'whether q' by a horizontal one. Note the difference between $\to$ and $\Rightarrow$.

(2) Assume $\mathcal{A} = \{p, q\}, W = \{pq, \bar{p}q, p\bar{q}, \bar{p}\bar{q}\}$. Then

$\mathbb{1}[p \Rightarrow ?q]$ is $\begin{array}{|c|c|}\hline pq & \bar{p}q \\ \hline p\bar{q} & \bar{p}\bar{q} \\ \hline \end{array}_W$, but $\mathbb{1}[?(p \to q)]$ would be $\begin{array}{|cc|}\hline \multicolumn{1}{|c}{pq} & \multicolumn{1}{c|}{\bar{p}q} \\ \hline \multicolumn{1}{|c|}{p\bar{q}} & \multicolumn{1}{c|}{\bar{p}\bar{q}} \\ \hline \end{array}_W$

[9] Using the standard equivalence $\phi \vee \psi \equiv \neg(\neg\phi \wedge \neg\psi)$, disjunction is no longer commutative either.

[10] Leaving out the restriction $\langle v, w\rangle \in \sigma[?\phi]$ it is no longer guaranteed that the result is an equivalence relation. Thanks to Gerhard Jäger for pointing this out.

[11] Combinations of updates behave *strict* with respect to undefinedness. So $\sigma[\neg\phi]$ or $\sigma[?\phi]$ undefined if either σ or $\sigma[\phi]$ undefined, and $\sigma[\phi \wedge \psi]$ and $\sigma[\phi \to \psi]$ undefined if either $\sigma, \sigma[\phi]$ or $\sigma[\phi][\psi]$ undefined.

2.2 Information Order, Support and Validity

In update semantics one usually finds a transitive and reflexive order $\sqsubseteq$ over information states, specifying information growth[12]. We say that $\sigma \sqsubseteq \tau$ iff τ contains at least as much information as σ. For formulas that do not cause undefinedness, the update function preserves information growth.

Definition 5 (Information Order i): $\sigma \sqsubseteq \tau$ iff $\sigma \supseteq \tau$

Fact 1 (Information Growth): For all σ and for all ϕ such that $\sigma[\phi]$ is defined, we have
1. $\mathbb{1} \sqsubseteq \sigma$, 2. $\sigma \sqsubseteq \mathbb{0}$, 3. $\sigma \sqsubseteq \sigma[\phi]$

Part 1. and 2. of fact 1 are true by definition: $\mathbb{1}$ contains no information, $\mathbb{0}$ trivially contains all information at once. Part 3. of fact 1 can be checked by induction on ϕ. The clauses for the atomic formulas and for negation are easily checked. (by definition of $\cap$) These can be called *declarative* updates. ($\sqsubseteq_d$) A declarative update increases information by eliminating possibilities only. For *interrogative* updates ($\sqsubseteq_i$), information growth is a little more subtle. Here information is increased by adding or refining issues. The partition induced by $\sigma[?\phi]$ is more fine-grained than that of σ. So fact 1 is satisfied for the interrogative dimension too. The clauses for $\wedge, \rightarrow, \Rightarrow$ and ∂ only combine other types of update. Definitions of declarative and interrogative updates can also be given directly.

Definition 6 (Information Growth ii):

$\sigma \sqsubseteq_d \tau$ iff $d(\sigma) \supseteq d(\tau)$
$\sigma \sqsubseteq_i \tau$ iff for all $\mathbf{v} \in i(\tau)$, there is a $\mathbf{w} \in i(\sigma)$ such that $\mathbf{v} \subseteq \mathbf{w}$

Once a notion of information growth is defined, it becomes possible to state the following key idea. An information state σ is said to contain ϕ, accept ϕ or *support* ϕ when updating σ with ϕ will not increase the information in σ.[13] Adding formulas that are already supported would be redundant.[14]

Definition 7 (Support):
$\sigma \Vdash \phi$ iff $\sigma[\phi] \sqsubseteq \sigma$

Update semantics employs a non-classical view on validity and entailment. An argument is valid if, after having applied all premises in the right order to some information state, the conclusion is supported. The second notion of entailment, $\models^{\mathbb{1}}$, represents entailment with a *closed world assumption*: if the premises are *all* you learned, then the conclusion is supported. A combination of presupposition accommodation and $\models^{\mathbb{1}}$ may produce non-monotonic inference. For instance $\partial p \models^{\mathbb{1}} p$ but $\neg p \wedge \partial p \not\models^{\mathbb{1}} p$.

[12] Obviously *real* information growth would be modeled by non-inclusive $\sqsubset$.
[13] By fact 1 $\sigma \sqsubseteq \sigma[\phi]$. Because $\sqsubseteq$ is anti-symmetric definition 7 becomes: $\sigma \Vdash \phi$ iff $\sigma[\phi] = \sigma$.
[14] Please note that this redundancy constraint heavily depends on the 'standard' rationality assumptions of epistemic logic. See footnote 5.

Definition 8 (Validity):

$\phi_1, ..., \phi_n \models \psi$ iff $\forall\sigma\ \sigma[\phi_1]...[\phi_n] \Vdash \psi$
$\phi_1, ..., \phi_n \models^{\mathbb{1}} \psi$ iff $\mathbb{1}[\phi_1]...[\phi_n] \Vdash \psi$

Having both declarative and interrogative updates is reminiscent of ideas proposed by C.S. Peirce (1867). According to Peirce information may be increased in two ways. By increasing the number of attributes used to describe a set of known objects, called comprehension or *depth*. Or by increasing the number of objects that are described, called extension or *breadth*. In the formal system of USQ depth and breadth can be defined. For each block in the partition, take the set of literals (atomic propositions or negated atomic propositions) that are shared by all possibilities in the block. The union of the atoms from these descriptions can be called the *current vocabulary* $\mathcal{A}(\sigma)$. Now depth corresponds to the set of predicates that occur in $\mathcal{A}(\sigma)$. Breadth corresponds to all referents, i.e. name-tags of objects, that occur in $\mathcal{A}(\sigma)$.

Definition 9 (Depth, Breadth):

$$\begin{array}{lcl} lits(\mathbf{v}) & = & \bigcap_{v\in\mathbf{v}}\{\ell \mid v(\ell) = 1,\ \ell \equiv Pa_1...a_n \text{ or } \ell \equiv \neg Pa_1...a_n\} \quad (\mathbf{v} \subseteq W) \\ \mathcal{A}(\sigma) & = & \{Pa_1...a_n \mid Pa_1...a_n \in lits(\mathbf{v}) \text{ or } \neg Pa_1...a_n \in lits(\mathbf{v}),\ \mathbf{v} \in i(\sigma)\} \\ depth(\sigma) & = & \{P^n \mid Pa_1...a_n \in \mathcal{A}(\sigma)\} \\ breadth(\sigma) & = & \{a_i \mid Pa_1...a_n \in \mathcal{A}(\sigma)\} \quad (1 \le i \le n) \end{array}$$

2.3 Resolving Issues

What remains to be specified, is what it means to answer a question or, in my terms, to *resolve an issue*. The notions of support and validity can be made to work for questions and issues as well. When a question does not add or refine any structure, there are two possibilities: (i) the structure induced by the question is already present in the information state. The issue is merely supported. (ii) The structure induced by the question only affects possibilities that are no longer part of the information state; the question is *answered* or the issue is *resolved* in that information state. (see lefthandside of figure 2) This can be expressed by taking $u(d(\sigma))$, which results in a partition of only one block. If that one block still supports the issue, the issue must be resolved.

Definition 10 (Resolution i): Question or issue $?\phi$ is fully resolved in σ when $u(d(\sigma)) \Vdash ?\phi$

New information ψ is said to be an *answer* to a question or to *resolve* an issue $?\phi$ in σ when σ updated with ψ resolves $?\phi$ in the earlier sense, but the old σ didn't.

Definition 11 (Resolution ii): Question or issue $?\phi$ is resolved by new information ψ in σ when $u(d(\sigma)) \nVdash ?\phi$ and $u(d(\sigma[\psi])) \Vdash ?\phi$

When there is information ψ that always answers a question $?\phi$, irrespective of the information state, we can say that that information constitutes a standard or *direct answer*. An *indirect answer* depends on the particular information state, for instance because it requires background knowledge.

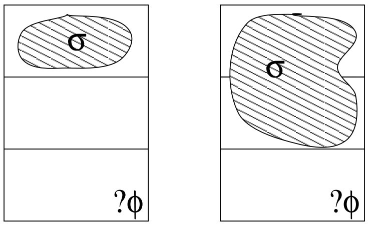


Fig. 2: full and partial resolution

Definition 12 (Direct Answer): New information ψ constitutes a direct answer to $?\phi$ when for all σ, $u(d(\sigma[\psi])) \Vdash ?\phi$

The one notion that remains to be captured is that of *partial resolution*. It is of crucial importance to the way in which relevance is explained in this paper. We say that information partially resolves an issue, when it eliminates all epistemic possibilities of at least one of the blocks in the partition. Taking out one of the alternative blocks, brings full resolution closer. A relevant utterance should provide information, that matters for the current issues. This definition captures precisely what it means for information to 'matter'. It must reduce the partition.

Definition 13 (Partial Resolution): Question or issue $?\phi$ is partially resolved by information ψ in σ when for at least one $\mathbf{w} \in i(\sigma[?\phi])$ there is no longer a $\mathbf{v} \in \sigma[?\phi][\psi]$ such that $\mathbf{v} \subseteq \mathbf{w}$.

The above definition can be made more elegantly using a non-inclusive interrogative information order $\sqsubset_i$. This information order captures *relevant information growth*.

Definition 14 (Relevant Information Growth): Information state τ contains more relevant information than σ, $\sigma \sqsubset_i \tau$, when for at least one $\mathbf{w} \in i(\sigma)$ there is no $\mathbf{v} \in i(\tau)$ such that $\mathbf{v} \subseteq \mathbf{w}$.

Please note that in this set-up answerhood and resolution always depend on the particular information state and on the current issue. In section 3 issues will be linked to user goals. So indirectly the notions of answerhood and resolution can be made dependent on user goals as well, when needed.

2.4 Quantification

There is one rather technical task that remains to be done. *USQ* needs to be extended for quantification. In this section I will only give a sketch of such an extension. The details have to be filled in later. For a proper account of questions and dynamic semantics see Groenendijk (1997).

Individual constants will function as referents or 'name-tags' of objects within the discourse.[15] For every referent a in the language, a possible object a^* is assumed.[16] For proper names like *John* a predicate $named(a_3, john)$ can be used. Quantification is dealt with by *substitution*. A substitution θ is a partial mapping from Var to $Refs$. Θ is the set of all substitutions. The substitution $\theta\{a/x\}$ is exactly like θ except $\theta(x) = a$. We say that $\theta \preceq \rho$ iff for all x such that $\theta(x)$ defined , $\theta(x) = \rho(x)$. The logical language L' is the normal language of predicate logic, extended with operators ∂ and ?. Terms t are either variables or referents. As in most dynamic semantics, the scope of an operator extends to the right. That is exactly the reason that quantification cannot be dealt with by substituting variables with referents in the formulas. Updates are thus transitions between states, subject to a substitution. The set $D(\sigma)$ is the current domain of σ. It contains all referents that are defined: $D(\sigma) = \{a \mid w(a) = 1 \text{ for all } w \in \sigma\}$.

Definition 15 (USQ'): USQ' is defined by $\langle \Sigma, [\phi]_\theta \rangle$ $(\phi \in L, \theta \in \Theta)$, where

$$\begin{array}{lcl}
\sigma[Pt_1...t_n]_\theta & = & \sigma \cap u(\{w \in d(\sigma) \mid w(P\theta(t_1)...\theta(t_n)) = 1\}) \text{ if } \theta(t_i) \neq undef \\
\sigma[\neg\phi]_\theta & = & \sigma \cap u(\{w \in d(\sigma) \mid w \notin d(\sigma[\phi]_\theta)\}) \\
\sigma[\phi \wedge \psi]_\theta & = & (\sigma[\phi]_\theta)[\psi]_{\theta'}\} \quad (\theta \preceq \theta') \\
\sigma[\phi \Rightarrow \psi]_\theta & = & \{\langle v, w\rangle \in \sigma[?\phi]_\theta \mid \text{ if } \langle v, w\rangle \in \sigma[\phi]_\theta, \text{ then } \langle v, w\rangle \in \sigma[\phi]_\theta[\psi]_{\theta'}\} \quad (\theta \preceq \theta') \\
\sigma[\exists x.\phi]_\theta & = & \sigma[\phi]_{\theta\{a/x\}} \quad (a \text{ fresh, i.e. } a \notin D(\sigma)) \\
\sigma[\forall x.\phi]_\theta & = & \bigcap_{a \in D(\sigma)} \sigma[\phi]_{\theta\{a/x\}} \\
\sigma[?\phi]_\theta & = & \{\langle v, w\rangle \in \sigma \mid v \in d(\sigma[\phi]_\theta) \text{ iff } w \in d(\sigma[\phi]_\theta)\} \\
\sigma[\partial\phi]_\theta & = & \sigma[\phi]_\theta, \text{ if } \sigma[\phi]_{\theta'} \neq \mathbb{0}, \text{ for some } \theta \preceq \theta' \\
 & & undefined, \text{ otherwise}
\end{array}$$

Special 'shortcut' operators are defined for expressing existential presupposition and wh-questions. (definition 16) The shortcut for existential presupposition is not defined as $\partial(\exists x.\phi)$ since that would always generate a fresh constant. Presupposed variables should unify with previously introduced referents. Only if those cannot be found, a fresh one is accommodated. The shortcut for wh-questions is not defined as $?(\forall x.\phi)$ since that would not give the partition predicted by Groenendijk and Stokhof (1984) where each block in the partition corresponds to an exhaustive answer. Obviously, one could argue for other ways of filling in these definitions. [17]

Definition 16 (Short Cuts): For all σ, ϕ

$$\begin{array}{lcl}
\sigma[\partial x.\phi]_\theta & = & \sigma[\phi]_{\theta'}, \text{ if } \sigma[\phi]_{\theta'} \neq \mathbb{0}, \text{ for some } \theta\{a/x\} \preceq \theta' \quad (a \in D(\sigma)) \text{ else} \\
 & & \sigma[\phi]_{\theta''}, \text{ if } \sigma[\phi]_{\theta''} \neq \mathbb{0}, \text{ for some } \theta\{b/x\} \preceq \theta'' \quad (b \notin D(\sigma)) \\
 & & undefined, \text{ otherwise} \\
\sigma[?x.\phi]_\theta & = & \bigcap_{A \subseteq D(\sigma)} \sigma[\bigwedge_{x_i \in X} ?\phi\{x_i/x\}]_{\theta\{A/X\}}
\end{array}$$

[15] In dynamic semantics one often finds three levels: *variables*, *referents* and *objects* from the domain. (Groenendijk et al. 1996) I only use variables and referents. In USQ there is no separate dimension for *discourse information*.

[16] Possibilities can be extended to predicate logic models $\langle D_w, I_w\rangle$, where $w(a) = 1$ iff $a^* \in D_w$ and $w(Pa_1...a_n) = 1$ iff $\langle a_1^*, ..., a_n^*\rangle \in I_w(P^n)$.

[17] Notation $\theta\{A/X\}$ is short for $\theta\{a_1/x_1, ..., a_n/x_n\}$, where X is a set of variables of the same size as A, possibly empty. $\phi\{x_i/x\}$ stands for ϕ with all free occurrences of x replaced by x_i.

3 Raising Issues

The characterization of relevance as proposed in the introduction crucially depends on something called 'question under discussion' or *current issue*. How are issues raised? Obviously issues may be raised by questions or other linguistic constructions, like focus. First I will explain how wh-questions and focus constructions are translated into L'. USQ makes it possible to deal with implicit domain restriction in a natural way (section 3.1). But issues can also be raised by non-linguistic means. In section 3.2 I will discuss the influence of a pre-defined topic-structure on issues. In section 3.3 it is explained how the AI repertoire of *actions*, *plans* and *goals* can be used alongside semantics, to predict possible issues in cooperative dialogue. In section 3.4 an example is given of actions from our SCHISMA dialogue system.

3.1 Questions, Focus and Domain Restriction

Issues can be raised by explicit questions. However, for wh-questions there are many ambiguities that become apparent in the formal model. A lot of choice remains with respect to translation of natural language constructs into L'. The basic idea is that wh-questions can be treated as most quantifiers in discourse semantics: by a conditional structure. The 'which-part' of the question, expressed by the antecedent of the conditional, restricts the relevant *context-set*. (cf. Westerståhl 1985)[18]

(3) Which Athenian is wise?

(4) a. $?x.(ath(x) \Rightarrow wise(x))$
b. $?x.(\partial y.(ath(y) \wedge wise(y)) \Rightarrow y = x)$

I suggest two translations: (4a) and (4b). The choice depends obviously on one's opinion on presuppositions of questions: does a wh-question presuppose the existence of a positive instance of the question?[19] If so, (4b) would be the correct translation. Answering question (3) with *nobody* would then 'cancel' the presupposition. This is possible on some views of presupposition, when the presupposition is triggered in the antecedent of a conditional. When one takes the view that a wh-question does not presuppose a positive instance, (4a) would be the right translation. In that case *nobody* is a perfectly good answer by itself.

I believe that the choice depends on the exact wording of the question and on its extra-linguistic setting. The fact that explicitly denying the presupposition makes sense, as in (5a), shows there is a case for the presupposition view. The fact that we often need a stronger form, as in (5b), shows that the presupposition is not obvious by itself. Background knowledge plays a decisive role. When we would replace 'Athenian' with 'Spartan', the non-presupposing reading of the question would be more appropriate.

[18] Example adapted from Jäger (1996), but note that Jäger paraphrases (3) as $\forall x.(ath(x) \Rightarrow ?wise(x))$ which gives different predictions.

[19] The example is a bit complicated by the fact that *which Athenians*, being a quantificational phrase, presupposes the existence of some Athenians anyway. This is not what is at stake here.

(5) a. Which Athenian, if any, is wise?
b. Which Athenian is the wise one?

Not all wh-questions have a 'which-part' that acts as a restriction. But normally there is *implicit domain restriction.* In example (6) the domain of discourse is implicitly restricted to those people that we know and that we would expect at such a party[20]. For instance, colleagues.

(6) Who was at the party?
$?z.(person(z) \wedge \partial x.party(x) \wedge likely\_to\_be\_at(x, z) \Rightarrow was\_at(x, z))$

Domain restriction has truth-conditional effects for focus-sensitive constructions, like negation and the particle *only.* Combining Rooth's (1992) semantics of focus with update semantics and questions makes it possible to explain this (Jäger 1996).

The *focus semantic value* of Rooth can be seen as an issue, namely the issue which one of the alternatives for the element in focus, is the actual one. Without domain restriction, the statement 'Only Socrates is wise.' would imply that for instance Zeus is not wise. When the issue in the dialogue was rather to find out 'Which Athenian is wise?', thereby implicitly restricting the domain of discourse to Athenians, no such dangerous implications would be possible.

Normally the focus semantic value is triggered by intonation contour, or some other marked focus construction. Assume a type-hierarchy for all objects. In example (7) the type of 'green' is *colour*, one node 'up' in the hierarchy.[21] The focus semantic value, would thus be the set of all colours. This corresponds to the alternatives in the partition induced by a question *What colour?*.

(7) The door is not $[\text{green}]_F$. $\mapsto$
"The door has a colour. What colour? It's not green."

Focus determines the scope of negation. There is an interesting connection here with a theory of *corrections.* A correction replaces the element in focus: the answer to the current issue.

(8) The door is $[\text{green}]_F$.
a. No, it's red.
b. No, it doesn't have a colour. It's a glass door.

In example (8a) the correction replaces the colour. But in (8b) it replaces the previously presupposed issue, that the door has some colour. So corrections may replace at various levels. It would be interesting to see how corrections could be dealt with in structured information states. More research is needed for that.

[20] Taking the non-presupposing reading here.
[21] In general, to derive the focus semantic value one needs *higher order unification* (Gardent and Kohlhase 1996).

We can conclude from this section that domain restriction plays an important role. Issues are able to represent the effects of explicit or implicit domain restriction. We have seen that issues can be raised by explicit questions, by focus constructions and by corrections. Issues are restricted by background knowledge.

3.2 Issues and Topics

Issues are meant to provide *discourse structure*. Different research traditions take different aspects of discourse as the central one. I follow Van Kuppevelt (1995) in taking the question dimension as the central one. Van Kuppevelt uses explicit and implicit questions to specify the *topic* of a sentence or larger part of discourse. Like questions or issues, the topic specifies what the discourse or sentence *is about*. Yet topics themselves are things, or objects. They are not abstract structuring elements. Topics do have a structuring effect, because each topic carries with it a *topic-space* of attributes and related topics. A good example of this structuring effect are the flight information dialogues studied by Rats (1996). A topic, often a flight, would be introduced by a description or by the flight number. Then questions about the attributes of the topic, such as time of arrival, would be asked and answered. Once established, the topic would be referred to using reduced expressions, like pronouns, or it would simply be omitted.

What is the influence of topic structure on the prediction of issues? Imagine an predefined data structure, that specifies, among other things, all potential topics in the task domain. For SCHISMA that would be performances, actors, titles, dates, prices etc. The data structure can be represented as a type hierarchy for feature-structures or object-oriented classes. There are *objects* of certain *types*. Objects have certain *attributes*, based on their type. Attributes may themselves refer to objects again; that would be potential sub-topics. During the conversation, the actual topic structure will therefore be constrained by the hierarchy. Topic structure is of great importance for anaphora resolution. Given the topic, antecedents are most likely to be found within the topic space. Nominal anaphors like *the wheel* are easy to resolve, when the topic is a car.

When a topic is introduced, not all of its attributes are known. For instance in example (1) the topic is a terrorist attack. Various features of the attack may lead to new issues. The possible follow-up issues are therefore constrained by the normal attributes of terrorist attacks in news stories. For narrative similar observations can be made. Imagine a story starting with: 'Once upon a time there was a little prince.' Now we get curious. Background knowledge about fairy-tales predicts issues like: What is his name? Is he blond? Are there any dragons around?

So by the introduction of a new topic, issues about the value of the attributes of the topic are raised. One might call these *curiosity issues*. Curiosity issues are constrained by the type hierarchy. For highly structured task domains, like travel information, the type-hierarchy can be much more accurate, than for low structured task domains, such as fairy tales.

3.3 User Goals

We have seen the influence on issues of linguistic expressions and of background knowledge in the form of type hierarchies. Often issues can also be raised by the situation we find ourselves in. To be more precise, issues are raised by the communicative goal that people normally have in such a situation. When I go up to a ticket-office in Budapest and utter a clumsy *harom jegyet*, I'll get my three tickets. The situation asks for a number. More words are not needed.

Ginzburg (1995) gives an number of examples to show that the apparent perlocutionary goal of the user is crucial. Yet in semantics one would like to abstract over such extra contextual parameters and establish *the* meaning of a linguistic construction. I suggest to make the distinction as follows. Since meaning in general is conceived of as the exchange of information, and since issues structure information, issues are part of meaning proper. Answerhood, and therefore the meaning of questions, is influenced by user goals. However, determining and predicting user goals is not itself part of semantics or pragmatics. It is part of AI and for dialogue systems it is part of the 'engineering' task. What we do need is a connection between goals, actions and issues, in such a way that once a set of user goals are given along with the possible actions to achieve them, the issues can be calculated.

At the moment I don't have a general principle to connect goals with issues. However, there is the concrete example of the SCHISMA dialogue system. (section 3.4) Moderate conclusions can be drown on the basis of the example.

There is a long tradition in AI of plans and goals to model human reasoning. (e.g. Allen et al. 1991) Often planning has been modeled as deduction. From an initial state the desired goal-state is to be deduced. *Actions* are steps of the plan. They correspond to deduction rules. Actions may consist of sub-actions or plans to achieve sub-goals. So we get a nice hierarchical structure. Heuristics may be used to calculate the estimated best plan. Actions and plans can be modeled using a version of dynamic logic (Harel 1984). States can be described using propositions or USQ ; actions by labeled transitions between states.

The following observation suggests a minimal theory of actions for simple inquiry dialogues. An action is a transition between states of the world. For simple inquiry dialogues, the world consists of information only. I believe, that inquiry dialogues are completely *information-driven*: all of the actions can ultimately be reduced to manipulating basic information. Therefore the hierarchical structure of possible actions and sub-actions will be isomorphic to the topic structure. For each attribute basic actions $request(attribute)$ and $inform(attribute)$ are defined. For pure information driven dialogues, all actions in the task domain can be build up from these basic actions.

3.4 A Dialogue System

In this section I will explain the running example of our SCHISMA system in more detail. The system provides the sort of information about performances you would normally find in a 'going-out' magazine. The system also allows people to reserve tickets. People who choose to use the system, will have a user-goal that corresponds with these capabilities. Based on a study of the theatre domain, a task model was developed with system *actions*. Each action corresponds to a a sequence of sub-actions, like querying the database, that will lead to a certain response. Actions have preconditions. For instance, a reservation requires a name, a definite performance and the number of tickets. The system is build to be cooperative. The system will plan actions to achieve the expected user goal. In deliberating the possible actions, preconditions of actions are tested. The system will ask for information elements, that are not already part of the context. So issues are raised about preconditions: the information elements that are required for an action. Here is an overview of the actions we have in the system. The information elements that are required are indicated.

show show and thereby select one or more performances
> requires either date, genre or artist
specific information gives information like director, actors, date, time, price or the number of available tickets
> requires (unique) selection
reserve make a reservation
> requires unique selection, name, number of tickets, discount type
< changes db: decreases available seats
esc escape from current action

SCHISMA is a so-called *mixed initiative* dialogue system. Initially the user has the initiative. The user asks questions. The systems responds by performing the appropriate action. But sometimes the system takes initiative. For a reservation a unique performance is required. When no performance is selected, the system will ask questions or make suggestions to help the user make up his mind.

So, we can conclude that new issues can be raised by the user asking actions of the system, or by the system acting cooperatively. In both cases the preconditions of actions raise new issues.

Sometimes the system makes a wrong guess. Too often it assumes that a reservation is wanted. When a user finds himself in an unwanted sub-dialogue or simply when the user changes plan, the current action can be cancelled, thereby also cancelling the corresponding issues. This is called the *escape* option, by analogy of programs where `Esc` is used to return to the program level just above.

So finally, we conclude that explicit protesting may change issues.

4 Summary and Future Research

The goal of the paper was to give a formal characterization of *relevance*. First, I have presented a logical system, USQ, that characterizes information change in a dialogue context. The system gives an account of information as structured data. Structure is provided by issues. An issue is modeled as a partition that determines which basic facts are relevant for the 'question under discussion'. An issue is resolved, when it no longer structures the data. An issue is partially resolved, when it reduces the number of distinctions made by the partition. The notions of 'resolving an issue' and 'direct answer' are straightforward adaptations of update semantics notions. So the first conclusion must be that, given a structure of issues, it possible to define a rather technical and reduced, notion of relevance.

Second, I have discussed several ways in which issues may be raised. The first three are more or less general 'linguistic' ways in which issues may be raised. The last three are more particular to inquiry systems like SCHISMA . Issues may be raised in the following ways:

(1) by a question.
(2) by a focus-expression. The *focus semantic object* becomes an issue.
(3) by a new topic. A new topic normally raises 'curiosity issues' about the attributes of the topic.
(4) by the user requesting an action of the system. Each action has a precondition: it requires attributes to be filled in. The value of these attributes will become an issue.
(5) by the system acting cooperatively. Actions can be planned to meet user goals. The precondition of such actions will become an issue.
(6) by the user explicitly protesting and indicating a change of issue.

However, structured information states have more benefits. The following observations may give food for thought.

I have discussed the use of a *type-hierarchy* at several points in this paper. It would be interesting to work out a typed version of USQ. Some of the proposals in this paper can then be tested. A typed version of USQ would be easier to implement on a computer. It would be interesting to find a computational counterpart of USQ : a way of storing the dialogue context, structured by issues. Each equivalence class in the partition can be described by a set of literals. (definition 9) So an information state can be described by a disjunction of conjunctions of literals. $repr(\sigma) = \bigvee_{\mathbf{v} \in i(\sigma)} \bigwedge lits(\mathbf{v})$. There is an interesting link here with SOAs in situation theory. See (Barwise and Perry 1983; Ginzburg 1996) and references cited there. Like literals, SOAs represent positive facts or negative facts. SOAs form a Heyting Algebra closed under meet ($\wedge$) and join ($\vee$). An actual situation can realize the facts described in a SOA. The equivalent of lambda-terms, so-called *SOA-abstracts*, are used by Ginzburg to specify $rel(?\phi)$, the relation that is commonly associated with a question. For instance rel('Who kisses whom?') $= \lambda x \lambda y.kiss(x, y)$. Adding lambda abstraction to the representation language, I expect that we would be able to reproduce $rel(?\phi)$ from a description of the information state $\mathbb{1}[?\phi]$.

Ginzburg calls $rel(?\phi)$ one of the three fundamental invariants needed to describe questions (1996;p 395). The other invariants are $aboutness\_set(?\phi)$, the set of propositions that constitute information *about* the question and $exh\_answer(?\phi)$, the proposition that, given a particular user goal, provides an account of when a question is resolved. The aboutness set corresponds, I think, to the set of *direct answers* as defined in definition 12. The exhaustive answer corresponds, I think, to the notion of an answer from definition 11.

Type hierarchies and the information order on issues $\sqsubseteq_i$ might be used to define an *epistemic entrenchment* order for AGM style *contraction* and *revision* (Gärdenfors 1988). Issues would make it is possible to deal with inconsistencies *locally*. Propositions that are 'under discussion' are less epistemically entrenched.

So, I would like to conclude that the idea of treating information states as being structured, has resolved some issues about the semantics of questions and answers, the semantics of focus, implicit domain restriction and the role of user goals in semantic theory. It has also raised some more issues, in particular about the possible role of a type hierarchy and about implementing structured information states on a computer.

References

Allen, J., Kautz, H., Pelavin, R., and Tenenberg, J.: 1991, *Reasoning about Plans*, Morgan Kaufmann Publishers, Inc., USA

Barwise, J. and Perry, J.: 1983, *Situations and Attitudes*, MIT Press

Beaver, D.: 1996, Presupposition, in J. Van Benthem and A. Ter Meulen (eds.), *Handbook of Logic and Language*, North-Holland, Elsevier

Gärdenfors, P.: 1988, *Knowledge in Flux: Modelling the Dynamics of Epistemic states*, Bradford Books

Gardent, C. and Kohlhase, M.: 1996, Focus and higher order unification, in *COLING-96*, Copenhagen, Denmark

Ginzburg, J.: 1995, Resolving questions, I, *Linguistics and Philosophy* 18, 459–527

Ginzburg, J.: 1996, Interrogatives: Questions, facts and dialogue, in S. Lappin (ed.), *The Handbook of Contemporary Semantic Theory*, Chapt. 15, Blackwell, Oxford

Grice, H. P.: 1975, Logic and conversation, in P. Cole and J. L. Morgan (eds.), *Syntax and Semantics*, Vol. 3, Academic Press

Groenendijk, J.: 1997, Questions in dynamic semantics, in G. Jäger and A. Benz (eds.), *Mundial '97*, University of Munich, this volume

Groenendijk, J. and Stokhof, M.: 1984, *Studies on the Semantics of Questions and the Pragmatics of Answers*, Ph.D. thesis, University of Amsterdam

Groenendijk, J. and Stokhof, M.: 1996, Questions, in J. Van Benthem and A. Ter Meulen (eds.), *Handbook of Logic and Language*, North-Holland, Elsevier

Groenendijk, J., Stokhof, M., and Veltman, F.: 1996, Coreference and modality, in S. Lappin (ed.), *Handbook of Contemporary Semantic Theory*, Blackwell, Oxford

Harel, D.: 1984, Dynamic logic, in D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, pp 497–604, Reidel, Dordrecht

Hoeven, G. F. v. d. et al.: 1995, Schisma a natural language accessible theatre information and booking system, in *First International Workshop on Applications of Natural Language to Data Bases*, Versailles, France

Hulstijn, J.: 1996, Presupposition accommodation in a constructive update semantics, in G. Durieux, W. Daelemans, and S. Gillis (eds.), *CLIN VI*, University of Antwerp, summary of Master's thesis, UvA, Amsterdam

Jäger, G.: 1996, Only updates - on the semantics of the focus particle *only.*, in *Proceedings of the 10th Amsterdam Colloquium*, ILLC, Amsterdam

Kuppevelt, J. v.: 1995, Discourse structure, topicality and questioning, *Journal of Linguistics* 31, 109–149

Peirce, C. S.: 1867, Upon logical comprehension and extension, in M. H. Fisch et al. (eds.), *Writings of Charles S. Peirce : a chronological edition*, Vol. I, 1857-1866, pp 70–86, Indiana University Press, Bloomington, 1982 edition

Rats, M.: 1996, *Topic Management in Information Dialogues*, Ph.D. thesis, Katholieke Universiteit Brabant, Tilburg

Rooth, M.: 1992, A theory of focus interpretation, *Natural Language Semantics* 1, 75–116

Stalnaker, R.: 1979, Assertion, in P. Cole (ed.), *Pragmatics*, No. 9 in Syntax and Semantics, pp 315–32, Academic Press, New York

Stalnaker, R.: 1984, *Inquiry*, MIT Press

Veltman, F.: 1996, Defaults in update semantics, *Journal of philosophical logic* 25(3), 221–262, previous version in DYANA R2.5A, 1990

Walker, M. A.: 1996, The effect of resource limits and task complexity on collaborative planning in dialogue, *Artificial Intelligence* 85, 181–243

Westerståhl, D.: 1985, Determiners and context sets, in J. van Benthem and A. ter Meulen (eds.), *Generalized Quantifiers in Natural Language*, Foris Publications, Dordrecht

Zeevat, H.: 1997, The Common Ground as a Dialogue Parameter, in G. Jäger and A. Benz (eds.), *Mundial '97*, University of Munich, this volume

The Construction of Answers*

Paul Piwek

IPO, Center for Research on User-System Interaction
P.O. Box 513, 5600 MB Eindhoven
The Netherlands, Email: piwek@ipo.tue.nl

1 Introduction

In the past fifteen years, context-dependence and context change have become central to the trade of formal semantics. In particular, it has become common ground that these notions are essential for the treatment of anaphora and presuppositions (see Kamp and Reyle 1993, Van der Sandt 1992). The importance of context for the notion of answerhood has been pointed out in Hintikka 1974.

In this paper, we propose a formalization of the notion of *answerhood* which treats answerhood as being essentially context-dependent. For this purpose, we take the informal definition of answerhood presented in Groenendijk and Stokhof 1984 as our point of departure:

> *A proposition gives an answer to a question in an information set, if the information set to which that proposition is added offers an answer.* Groenendijk and Stokhof 1984, page 154

The framework which we employ for the formalization of this definition is sufficiently general to cover answerhood and also the other aforementioned context-dependent phenomena, i.e., anaphora and presuppositions (see, e.g., Krahmer and Piwek 1997). The framework is based on a class of mathematical formalisms known as Constructive Type Theories (CTT; see, for instance, Barendregt 1992, De Bruijn 1980, Martin-Löf 1984).

CTT is suitable for dealing with context-dependent phenomena because it can be seen as a higher order generalization of Kamp's Discourse Representation Theory (Kamp 1981), henceforth DRT (see Ahn and Kolb 1990). For our purposes, the main advantage of CTT over DRT is that CTT has a well-studied proof system. This enables us to give a deductive account of answerhood. A direct benefit of a deductive account is that it can serve as the basis of a computational model.

* Thanks are due to the participants of the CLIN VII meeting in Eindhoven (November, 1996) and the MunDial workshop in München (March, 1997) for stimulating remarks and comments. Furthermore, I benefitted from discussions with René Ahn, Robbert-Jan Beun and Tijn Borghuis on an earlier version of this paper. The author is partially funded by the co-operation unit of Brabant Universities (SOBU).

On the conceptual side, our approach presents a bridge between approaches to answerhood which take context into account and a class of context-independent accounts which deal with questions as structures with gaps (or variables) and answers as the objects that can fill such gaps (Cohen 1929, Jespersen 1933, Katz 1968, Scha 1983, Prüst et al. 1994).

This paper leads up to an analysis of indirect answers (for instance, 'You may go, if you have finished your homework' is an indirect answer to the question 'May I go now?'). We discern a subclass of indirect answers, called conditional answers, which have a nice property. Furthermore, we give a formalization of the notion of conversational equivalence, which is essential for the proper treatment of indirect answers. Groenendijk and Stokhof 1984 stopped short of such a formalization. Finally, we introduce a novel type of answer, namely the preventive answer.

The paper is structured as follows. In section 2, we present some reasons for taking context into account in any definition of the notion of answerhood. In section 3, an informal outline of CTT is presented. In section 4, we sketch how questions and answers can be represented in the CTT language. In section 5, a formalization of the definition of answerhood is given. Furthermore, special attention is paid to the notion of indirect answerhood and we outline how the formalization can give rise to a computational model of answerhood. Finally, our conclusions are listed in section 6.

2 The Role of Context

In this section, we present four general grounds for rejecting any definition of answerhood which does not take context-dependence seriously.

First, whether a question can be posed depends on the background. There seems no point in asking a question whose answer is already part of common background of the interlocutors (keeping rhetorical questions aside). For instance, if the interlocutors share the information that *nobody has seen Mary*, then the question 'Who has seen Mary?' is inappropriate.[1]

Second, there are answers that only present a filler to the gap of a question in a certain context. For example, in a situation in which John asks 'Where is Mary?' and it is part of the context (more specifically, the common background) that *If her car is in the garage, then Mary is at home*, John's question can be answered with the sentence 'Her car is in the garage'. This answer does not, on its own, present the filler for the gap which belongs to the wh-constituent 'where'.

Third, some answers rule out certain fillers, rather than presenting them. This sort of answer will be called a negative answer. For example, 'Where is Mary?' can be answered

[1] It is often assumed that wh-questions carry existential presuppositions. For instance, 'Who walks?' is said to presuppose that *somebody walks*. A consequence of this claim is that 'Nobody walks' does not count as a regular answer to the aforementioned question. We agree with Groenendijk and Stokhof 1997 that this result is not desirable. Rather we think that the relevant constraint that a question for information puts on the context, i.e., the common background, is that the question is not already answered in the common background. For more on this issue, see Piwek 1997.

with the sentence ‘She is not at home’. Here context change plays an important role: the answer changes the context into one in which ‘At home’ is no longer a possible filler.

Fourth, there are answers which raise new questions whose answers bring the questioner closer to an answer for the original question. Such answers are called *indirect answers*. Consider ‘Mary is at home if her car is in the garage’ as an answer to the question ‘Where is Mary?’. Groenendijk and Stokhof 1984 point out that the answer provides the questioner with a new question (‘Is Mary’s car in the garage?’), which if answered positively also resolves the original question.

3 Constructive Type Theory

In this section, CTT is described by comparing it with DRT (cf. Ahn and Kolb 1990, Ranta 1990). In DRT, contexts are modelled as Discourse Representation Structures (DRSs). A DRS consists of a set of discourse referents and a set of conditions. The discourse referents can be seen as pegs, and the conditions as assignments of properties to these pegs. For instance, if the sentence *A horse neighs* is processed in an empty context, then a fresh referent and two conditions are added to the empty context. This yields the new context $[x \mid horse(x),\ neigh(x)]$ (the empty context is represented as follows: $[\,|\,]$). The conditions attach the properties *is a horse* and *neighs* to the referent.

3.1 Introductions

In CTT, a context is modelled as a sequence of introductions. Introductions are of the form $V : T$, where V is a variable and T is the type of the variable. A context is an ordered sequence of introductions, because the type T of an introduction may depend on other introductions that precede it. We give an example of this shortly.

The variable V in an introduction $V : T$, where T itself is of the type *set* (i.e., $T : set$), corresponds to a discourse referent in DRT. For instance, a referent for an entity from the set of horses is introduced as follows: $x : horse$. The type *horse* should only be used in the introduction $x : horse$ if $horse : set$ is already part of the context. In other words, the introduction $horse : set$ has to precede the introduction $x : horse$. This way, one introduction can depend on another introduction. Of course, the same argument applies to the type *set*. This type is an inhabitant of the box ($\Box$), which by definition does not have a type.

DRT’s conditions correspond to introductions $V : T$, where T is of the type *prop* (short for proposition, again we have $prop : \Box$). Thus, the introduction $y : neigh \cdot x$ corresponds to the condition *neigh(x)*. The type $neigh \cdot x$ (of type *prop*) is obtained by applying the type *neigh* to the object x. Therefore, it depends on the introductions of x and *neigh*. Since x is of type *horse* and $neigh \cdot x$ should be of the type *prop*, *neigh* must be a (function) type from the set of horses into propositions, i.e., $neigh : horse \rightarrow prop$.

The introduction $y : neigh \cdot x$ involves the variable y (of the type $neigh \cdot x$). The variable y is said to be an inhabitant of $neigh \cdot x$. This raises the question what sort of an entity the inhabitant of a proposition could possibly be. Curry and Feys 1958 observed that propositions can be seen as classifying proofs (this is known as the 'propositions as types – proofs as objects' interpretation of Type Theory). This means that the aforementioned introduction states that there is a proof y for the proposition $neigh \cdot x$.[2]

3.2 Function types

In DRT, the proposition *Every horse neighs* is translated into the implicative condition:

$$[x \mid horse(x)] \Rightarrow [\,\mid neigh(x)].$$

In CTT this proposition corresponds to the type $(\Pi x : horse.neigh \cdot x)$, which is a dependent function type. It describes functions from the type *horse* into the type $neigh \cdot x$. The range of such a function $(neigh \cdot x)$ depends on the object x to which it is applied. Suppose that we have an inhabitant f of this function type, i.e., $f : (\Pi x : horse.neigh \cdot x)$. Then we have a function which, when it is applied to an arbitrary object y of type *horse*, yields an inhabitant $(f \cdot y)$ of the type $neigh \cdot y$. In other words, f is a constructive proof for the proposition that *Every horse neighs*.

3.3 Deduction

CTT encompasses a number of deduction rules with which one can determine the type of an object in a given context. These rules can also be used to search for an object belonging to a particular type. In other words, the rules enable us to check whether in a context Γ it can be derived that object E is of type T (notation: $\Gamma \vdash E : T$).

Here, not all rules can be discussed. We limit our attention to two rules which are relevant in the context of this paper. The first rule is rather straightforward; it says that if we have a context that contains the introduction $V : T$, then we can deduce that in that context V is of type T (i.e., $V : T$). Below, in the representation of the rule, $\otimes$ stands for the concatenation of sequences. Γ_1 and Γ_2 are sequences.

$$\frac{(\Gamma_1 \otimes V\!:\!T \otimes \Gamma_2) \text{ is a context}}{(\Gamma_1 \otimes V\!:\!T \otimes \Gamma_2) \vdash V\!:\!T}\ \text{(selection)}$$

Let us sketch a situation in which this rule can be employed. Suppose we have some context Γ containing the introduction $p : neigh \cdot r$. Furthermore, assume that we want to find out whether an inhabitant of (in other words, a proof for) $neigh \cdot r$ can be derived in Γ. This problem can be stated as follows: We are in search of a substitution $\langle S \rangle$ such that:

(1) $\Gamma \vdash X : neigh \cdot r\langle S \rangle$

[2] For the readers who are more accustomed with possible-worlds semantics, compare the definition of a proposition as a set of possible worlds, with the idea that a proposition corresponds to the set of its proofs. Both these implementations of the notion of a proposition are in line with the intuition that somebody grasps a proposition if he or she knows in which situations it holds or doesn't hold. (Replace situation with world/proof, respectively.)

In (1), the capital X is a so-called *gap*. The task is to find a substitution $\langle S\rangle$ for this gap such that (1) can be derived. A substitution is a list of assignments of CTT expressions to gaps, e.g., $\langle G_1 := E_1, \ldots, G_n := E_n\rangle$ (where $G_1, \ldots, G_n$ are gaps and $E_1, \ldots, E_n$ are CTT expressions). (1) can indeed be deduced if we assume that $\langle S\rangle$ is equal to $\langle X := p\rangle$, because in that case we can use the selection rule. We have an instance of $(\Gamma_1 \otimes V : T \otimes \Gamma_2) \vdash V : T$ (namely, $\Gamma \vdash p : \mathit{neigh} \cdot r$; remember that $p : \mathit{neigh} \cdot r$ is a member of Γ) and thus, according to the deduction rule, have to check whether Γ is a context. This is indeed so, since we assumed it at the outset.

With a proof system which is limited to the selection rule, one can only check whether an object is of a particular type by determining whether an introduction to this effect is a member of the context. CTT, however, has more deduction rules. One of them allows us to combine the information of different introductions. This rule bears a close resemblance to the Modus Ponens scheme of Propositional Logic, and also involves function application:

$$\frac{\Gamma \vdash F : (\Pi x : A.B) \quad \Gamma \vdash a : A}{\Gamma \vdash F \cdot a : B\langle x := a\rangle} \text{ (application)}$$

For instance, if a context Γ contains the introductions $g : (\Pi y : \mathit{horse}.\mathit{neigh} \cdot y)$ and $b : \mathit{horse}$, then we can use this rule to find an inhabitant of the type $\mathit{neigh} \cdot b$. In other words, our goal is to find a substitution $\langle S\rangle$ such that $\Gamma \vdash P : \mathit{neigh} \cdot b\langle S\rangle$. The substitution $\langle S\rangle$ should assign a value to the gap P.

The deduction rule tells us that $(g \cdot b)$ can be substituted for P, if $\Gamma \vdash g : (\Pi y : \mathit{horse}.\mathit{neigh} \cdot y)$ and $\Gamma \vdash b : \mathit{horse}$. Both can be deduced using the selection rule, because we assumed that $g : (\Pi y : \mathit{horse}.\mathit{neigh} \cdot y)$ and $b : \mathit{horse}$ are members of Γ. Thus, we conclude that $\Gamma \vdash (g \cdot b) : \mathit{neigh} \cdot b$.

Below we use $\vdash_\Delta$ for the deduction of more than one conclusion in a context. $\vdash_\Delta$ is defined as follows:

D. 1: $\Gamma \vdash_\Delta S_1, \ldots, S_n$ if $\Gamma \vdash S_1, \ldots, \Gamma \vdash S_n$.

4 Representing questions and answers in CTT

Traditionally, questions are divided into three categories: (1) Yes/no-questions such as *Does John walk?*, (2) choice questions such as *Does John walk or does Mary walk?* and (3) Wh-questions, such as *Who entered the room?*.

Our analysis of wh-questions builds on the insight that a question is a structure which contains one or more well-defined gaps. This observation (first made by Cohen 1929 according to Bäuerle and Zimmermann 1991) draws upon the similarity between wh-questions and algebraic equations, which both feature some unknown quan tity. For instance, the variable x in the equation $x = 1 + 1$ seems to fulfil the same function as the wh-constituent 'what (number)' in 'What (number) is the sum of 1 and 1?' Both the variable x and the word *what* mark a gap in the formula/sentence. This gap can be filled with a value. A value is said to be a solution or an answer if filling the gap with the value produces a true formula or sentence.

We already saw that gaps also have a role in CTT. In section 3.3 above, we show how a substitution for a gap can be found with respect to a context Γ by using the deduction rules backwards. The idea is now to use this technique to check whether a context provides an answer to a question (this is the central concept in Groenendijk and Stokhof's definition of the question-answer relation).[3]

Note that as a consequence of this approach, the basic kind of answer (i.e., the filler of a gap) is not exhaustive. In other approaches, notably Groenendijk and Stokhof 1984, the basic type of answer is taken to be (strongly) exhaustive. For instance, a basic answer to 'who walks' is taken to be a proposition which (1) states for exactly those persons that walk, that they walk and (2) states that no one else walks (the closure condition).[4]

Before we illustrate our approach with an example, let us introduce two useful definitions:

D. 2: *A segment* is a sequence of introductions.

D. 3: *An open segment* is a sequence of introductions with at least one gap.

Take the question *Who walks?* We assume that there is some interpretation function which translates this question (given a context of interpretation) into the open segment: $X : \mathit{person}, P : \mathit{walk} \cdot X$. Furthermore, suppose that the introductions $\mathit{john} : \mathit{person}, p : \mathit{walk} \cdot \mathit{john}$ are a part of some context Γ. In that case, the question is answered in Γ. In other words, there is a substitution $\langle S \rangle$ such that:[5]

$$\Gamma \vdash_{\Delta} X : \mathit{person}, P : \mathit{walk} \cdot X \langle S \rangle$$

The substitution in question is $\langle X := \mathit{john}, P := p \rangle$. This substitution fills the gaps that occur in the representation of the question. Notice, that there is a difference between the X and the P gap. The former is the gap whose value the questioner is interested in: it is filled by $\langle S \rangle$ with a referent for a person. P is a gap which is filled with a proof that this person walks. The questioner is not interested in the identity of this proof: he is satisfied if he knows that there is a proof. Henceforth, we will write the former type of gaps in bold face and call them *marked* gaps to distinguish them from the second type of gaps.Thus the representation of *Who walks?* is $\mathbf{X} : \mathit{person}, P : \mathit{walk} \cdot \mathbf{X}$.

Yes/no-questions can be seen as a special kind of wh-question. For instance, the question *Does Mary walk?* corresponds to the following structure: *F is equal to (It is true that) or F is equal to (It is false that), and F Mary walks.* This interpretation corresponds to the following open segment:[6]

[3] Our work is related to that of Ahn 1994. He sketches the use gaps to represent wh-questions in CTT. His work is different from ours in that he does not give a (formal) definition of the notion of answerhood.

[4] In terms of our approach, exhaustiveness can be reconstructed by requiring that an answer presents *all* 'true' fillers for the gap in a question and a proposition expressing that any other fillers lead to a contradiction.

[5] $\vdash_{\Delta}$ is defined in definition 1.

[6] We assume that equal denotes β-equality. If two terms are β-equal, then they reduce to the same normal form.

$\mathbf{F} : prop \to prop,$
$G : equal(\mathbf{F}, \lambda x : prop.x) \vee equal(\mathbf{F}, \lambda x : prop.x \to \bot),$
$Q : \mathbf{F} \cdot (walk \cdot m)$

The translation of choice questions can be carried out along the same lines using the selection functions (e.g., $\lambda x.\lambda y.x$ and $\lambda x.\lambda y.y$). [7]

We have shown how the three basic types of questions can be represented in CTT. Let us now turn to the representation of answers.

We assume that answers to questions are propositions. It is beyond the scope of this paper to define a full mapping from natural language to propositions in CTT, but we do want to discuss some relevant issues. For instance, we assume that a full proposition can also be recovered from elliptical sentences, which often function as answers to questions.

If we construct a formal representation of a sentence, then a noun phrase is mapped to a gap. This gap is basically an underspecified representation of an object. In case of a definite noun phrase, the gap needs to be filled with a particular object from the context (see Krahmer and Piwek 1997).

If an indefinite occurs in a sentence, then the gap which corresponds to it does not have to be filled with a particular object from the context. The indefinite noun phrase stands for an indefinite/arbitrary object. Thus the gap remains present in the representation of the sentence.

5 Formalizing answerhood

Groenendijk and Stokhof 1984 formalized their definition of answerhood in possible-world semantics. They model an information set of an agent as a set of possible worlds, i.e., those worlds which are compatible with the information available to the agent. A question is taken to be a partition on this set (see also Hamblin 1971 for the use of partitions to model questions). For instance, a yes/no-question such as 'Does John walk?', partitions the information set I into a subset I_1 of worlds in which *John walks* and a subset I_2 of worlds in which *he doesn't walk*. A proposition P now counts as an answer if I' (I after it has been updated with P, i.e., all worlds which are incompatible with P are thrown out of I) is a subset of I_1 or of I_2. In other words, either *John walks* or *John doesn't walk* holds in the new information set.

This sketch of Groenendijk and Stokhof's formalization allows us to compare their formalization of the notion of answerhood with ours. Before we present the comparison, we give a more detailed picture of the kind of contexts with respect to which answerhood is defined in our theory and we spell out our formalization of answerhood in detail.

[7] For instance, the translation of *Does John walk or does Mary walk?* is
$\mathbf{F} : prop \to (prop \to prop),$
$G : (equal(\mathbf{F}, \lambda x : prop.\lambda y : prop.x) \vee equal(\mathbf{F}, \lambda x : prop.\lambda y : prop.y)),$
$Q : (\mathbf{F} \cdot (walk \cdot j)) \cdot (walk \cdot m)$

5.1 The common background as context

Below, the context (henceforth, Γ) is taken to correspond to the information which the questioner assumes to be the common background of the interlocutors. The notion of a common background is borrowed from Stalnaker 1974. Stalnaker proposes that the common background is the place where the presuppositions of a discourse are stored.

The fact that a dialogue participant assumes that certain information is commonly agreed upon, means that he *expects* none of the dialogue participants to act in a way which is incompatible with this information. Here we will assume, also following Stalnaker, that this means at least the following two things: a dialogue participant does not present information which contradicts the common background and also does not present information which already follows from the common background. The latter rule forces the dialogue participants to be informative with respect to the common background.

The notions of consistency and informativity have their formal counterparts in CTT:

D. 4 (Consistency): A segment A is consistent with respect to Γ iff it is not the case that there is a proof p such that $\Gamma \otimes A \vdash p : \bot$

In words, consistency of some segment A with a context Γ means that in Γ extended with A it is not possible to derive a proof for $\bot$.

D. 5 (Informativity): A segment A is informative with respect to Γ iff it is not the case that $\Gamma \vdash_\Delta A$

Thus, a segment is only informative with respect to Γ if it cannot be derived from Γ.

5.2 Positive answers and inference

We start by giving a definition of the notion of a positive answer.

D. 6 (Positive answer): A segment A is a positive answer for an open segment Q in a context Γ, iff there is a substitution S such that:

1. $\Gamma \otimes A \vdash_\Delta Q\langle S\rangle$;
2. $\Gamma \otimes A$ is consistent; and
3. For all $\langle S'\rangle$ which differ at most from $\langle S\rangle$ in their assignments to non-marked gaps, $Q\langle S'\rangle$ is informative with respect to Γ.

The definition says that A positively answers Q if there is a substitution $\langle S\rangle$, such that $Q\langle S\rangle$ is derivable in the context extended with the answer A. We use the term 'positive answer' to express that such an answer presents a specific substitution instance to the question at stake.

Besides condition *1.*, the conditions *2.* and *3.* on consistency and informativity have to be satisfied. The condition on consistency speaks for itself. The condition on informativity requires some explanatory remarks. Consider, for example, the substitution $\langle \mathbf{X} := john, P := q\rangle$ for the open segment (2).

(2) $\mathbf{X} : \mathit{person}, P : \mathit{walk} \cdot \mathbf{X}$

This open segment corresponds to the question in (3).

(3) Who walks?

Suppose we have an answer to this question which if it is added to the context Γ yields a substitution $\langle S' \rangle$. Suppose $\langle S' \rangle$ assigns the variable *john* to the gap **X** and some proof of the proposition *that John walks* to the gap P. Now, condition 3. on informativity ensures that the information expressed by the aforementioned answer (i.e., *that John walks*) is not already derivable in the common background Γ. The condition says that any segment which can be obtained from (2) by applying $\mathbf{X}{:=}\ \mathit{john}$, and some arbitrary substitution on P (in other words, a substitution which differs at most from $\langle S' \rangle$ in the variables that it assigns to non-marked gaps, which the questioner is not interested in), should be informative with respect to the context Γ. In other words, in Γ there should be no proof of the proposition that John walks.

Let us now have a closer look at condition *1.* Consider again the question 'Who walks?'. According to our definition 'John walks' is a positive answer to (3). The formal representation for this answer is $q : \mathit{walk} \cdot \mathit{john}$. If this introduction is added to the context Γ, then there is a substitution $\langle S \rangle$ such that (we assume that also the conditions *2.* and *3.* are fulfilled):

(4) $\Gamma \otimes q : \mathit{walk} \cdot \mathit{john} \vdash_{\Delta} \mathbf{X} : \mathit{person}, P : \mathit{walk} \cdot \mathbf{X}\langle S \rangle$

The substitution in question is $\langle \mathbf{X} := \mathit{john}, P := q \rangle$. In this case, there is a straightforward relation between the representations of the question and the answer: $q : \mathit{walk} \cdot \mathit{john}$ and $\mathbf{X} : \mathit{person}, P : \mathit{walk} \cdot \mathbf{X}$ can be unified by applying the aforementioned substitution (and, in fact, this is what the proof system does in order to verify the derivation; the selection rule is applied in this case, see section 3.3). This reveals the relation between our approach and the work by Katz 1968 up to the work by Prüst et al. 1994 in which unification is put forward as the basic mechanism for explaining the relation between questions and answers.

Unification is the basic operation of a simple proof system which consists of the selection rule. Let us now indicate the limits of such simple proof systems. Consider example (5) as a reply to question (3).

(5) John's car is broken.

We represent this sentence as follows, assuming that 'John's car' is represented in the context by the variable $\mathit{john's\_\ car}$:

(6) $r : \mathit{broken} \cdot \mathit{john's\_car}$

Evidently, no unification is possible between (2) and (6). Thus, unification rules (5) out as a positive answer. Now, suppose that it is common background that *if John's car is broken, then he walks*:

(7) $f : (\Pi a : (broken \cdot john's\_car).walk \cdot john)$

Given such a Γ, we would like to predict that (5) counts as a positive answer. This is precisely what can be done by employing a deductive system such as CTT. If we take $\langle S \rangle$ = $\langle$ **X**:=*john*, *P*:= $f \cdot r$ $\rangle$ then it holds that

(8) $\Gamma \otimes$ (6) $\vdash_{\Delta}$ (2) $\langle S \rangle$

In this case, a substitution for P is obtained by combining the background information and the answer. This is witnessed by the proof object which is assigned to P: this proof object consist of r (the proof object introduced by the answer) and f (the proof object for the conditional that is a member of the common background Γ).

5.3 Negative answers

Groenendijk and Stokhof 1984, who take exhaustive answers to be primary the type of answers, introduce partial answers as answers which rule out one or more exhaustive answers.

Analogously, we define a negative answer as an answer which rules out one or more positive answers to a question. For instance, 'John doesn't walk' is a negative answer to question (3), since it rules out the answer 'John walks'. The definition of a negative answer is given below. It contains the operator *NOT* which is first defined.

D. 7 (NOT): $NOT(\ldots, A : B)$ is equal to $\ldots, A : B \rightarrow \bot$.

The operator *NOT* transforms the main proposition of a segment into its negation. Let us illustrate the use of *NOT* with an example. (9.a) is an abbreviation of (9.b):

(9) a. $NOT(\mathbf{X} : person, P : walk \cdot \mathbf{X})$
    b. $\mathbf{X} : person, P : (walk \cdot \mathbf{X}) \rightarrow \bot$

D. 8 (Negative answer): A segment A is a negative answer for an open segment Q in a context Γ, iff there is a substitution S such that:

1. $\Gamma \otimes A \vdash_{\Delta} NOT(Q)\langle S \rangle$;
2. $\Gamma \otimes A$ is consistent; and
3. For all $\langle S' \rangle$ which differ at most from $\langle S \rangle$ in their assignments to non-marked gaps, $NOT(Q)\langle S' \rangle$ is informative with respect to Γ.

Thus, if in context Γ the question (3) corresponding to the representation (2) is posed, then we have a negative answer for this question if a substitution can be found for (9.b) in Γ extended with the answer. Suppose the answer is *John doesn't walk*. Formally, this answer is represented as $n : (walk \cdot john) \rightarrow \bot$. Now the substitution $\langle$**X**:=*john*, *P*:=*n*$\rangle$ can be found.

5.4 Indirect answers

Groenendijk and Stokhof 1984 think of indirect answers as providing the questioner with new ways for getting an answer to the original question. For instance, if the questioner asks whether ϕ, then *if* ψ *then* ϕ provides an indirect answer, because now the questioner can obtain an answer to the original question by finding out whether ψ holds. The indirect answer suggests a new question (i.e., whether ψ). If the questioner discovers that ψ holds, then the original question has automatically been answered.

Groenendijk and Stokhof formalize the notion of indirect answer roughly as follows: Given an information set I and a question Q, A is an indirect answer iff there is some question R in I' (I updated with A), such that more answers to R are (partial) answers to Q in I' than in I. Furthermore, the following condition has to hold: R and Q should not be conversationally equivalent.

Notice that this definition is computationally rather impractical. First, it requires coming up with a new question. There is nothing which guides us in the search for such a question. Second, there is the notion of conversational equivalence. Two questions are conversationally equivalent if 'the questioner has to assume that an informer will be able to answer the one question truthfully iff she is able to answer the other truthfully as well. So, if a proposition gives rise to a new question which is conversationally equivalent to the original one, the entire point of providing an indirect answer vanishes' (Groenendijk and Stokhof 1984, page 164). The problem is that Groenendijk and Stokhof do not provide a formal definition of conversational equivalence, whereas it does play an important role, as they themselves illustrate by the following thought experiment.

Suppose we have an information set with respect to which the following two atomic propositions ϕ and ψ are totally independent. Now it is impossible that *that* ϕ provides an indirect answer to the question *whether* ψ. It does so, however, if we do not take conversational equivalence into account: if ϕ is added to the information set, then *whether* ψ does depend more on the question *whether if* ϕ *then* ψ (since a positive answer to the question also provides an answer to the original question, whereas it didn't in the information set to which ϕ had not yet been added).

We show that the aforementioned two problems can be solved in our approach. The solution for the latter problem which we present in our framework can, however, also be translated to the framework of Groenendijk and Stokhof. Things are different with respect to the former problem, i.e., how to determine which new question becomes interesting after an indirect answer. In a syntactic approach this can be read of from the structure of the representation of the answer in case of a *conditional* answer. A conditional answer is a specific sort of indirect answer. For instance, the aforementioned indirect answer *if* ψ *then* ϕ to the question *whether* ϕ is a conditional answer. It translates into $f : (\Pi p : \psi.\phi)$. The relevant information is $p : \psi$, i.e., the (abstraction) domain of the Π-type.

Indirect answerhood can be tested as follows: add not only the answer $f : (\Pi p : \psi.\phi)$ to the context but also, temporarily, the relevant information $p : \psi$. Subsequently, check whether in this context $(\Gamma \otimes f : (\Pi p : \psi.\phi) \otimes p : \psi)$ the question is (positively/negatively) answered. In this case, the question can indeed be answered in the thus extended context;

a proof can be constructed for ψ (i.e., $f \cdot p$). In other words, there is an $\langle S \rangle$ such that in the context it can be derived that:[8]

$$F : prop \rightarrow prop,$$
$$G : equal(F, \lambda x : prop.x) \vee equal(F, \lambda x : prop.x \rightarrow \bot),$$
$$Q : F \cdot \psi \langle S \rangle$$

Let us now provide a formalization for the notion of conditional answerhood. For that purpose, we first have to define a function for obtaining the abstraction domain of a Π-type, such that it can be added to Γ. We do, however, have to be careful. Γ should not become inconsistent in the process of doing so. Thus, the function returns only those introductions which do not yield an inconsistent context. Formally, the segment consisting of the introduction of the abstraction domain of the conditional answer which can be added to the context is returned by the function Φ. This function is defined after the following notational convention, which makes reading and manipulating Π-types somewhat easier.

N.C. 1: $(\Pi\langle x_1 : T_1, \ldots, x_n : T_n \rangle.B)$ is an abbreviation for the Π-type $(\Pi x_1 : T_1. \ldots (\Pi x_n : T_n.B) \ldots)$.

D. 9 (Φ-segment): Given a statement $A = f : (\Pi\langle x_1 : T_1, \ldots, x_n : T_n \rangle.B)$ and a context Γ,

$\Phi(A) = x_1 : T_1, \ldots, x_m : T_m$ for $m \leq n$ if

1. $\Gamma \otimes x_1 : T_1, \ldots, x_m : T_m$ is consistent; and
2. (a) $m = n$ or
   (b) $\Gamma \otimes x_1 : T_1, \ldots \otimes x_{m+1} : T_{m+1}$ is inconsistent.

With the help of Φ, we can now define conditional answerhood.

D. 10 (Conditional answer): A statement A $(= f : \Pi\langle x_1 : T_1, \ldots, x_n : T_n \rangle.B)$ is a conditional answer for an open segment Q in a context Γ, iff there is a substitution S such that:

1. $\Gamma \otimes \Phi(A) \otimes A \vdash_\Delta Q\langle S \rangle$;
2. (consistency) A is consistent with respect to $\Gamma \otimes \Phi(A)$;
3. (informativity) For all $\langle S' \rangle$, $Q\langle S' \rangle$ is informative with respect to $\Gamma \otimes \Phi(A)$, where $\langle S' \rangle$ differs at most from $\langle S \rangle$ in the assignment to unmarked gaps.

or

1. $\Gamma \otimes \Phi(A) \otimes A \vdash_\Delta NOT(Q)\langle S \rangle$;

[8] The substitution in question is $\langle F := \lambda x : prop.x, Q := f \cdot p, G := r \rangle$, where r is a proof of the fact that $\lambda x : prop.x$ is equal to $\lambda x : prop.x$. This follows from the reflexivity axiom for equality.

2. (consistency) A is consistent with respect to $\Gamma \otimes \Phi(A)$;

3. (informativity) For all $\langle S' \rangle$, $NOT(Q)\langle S' \rangle$ is informative with respect to $\Gamma \otimes \Phi(A)$, where $\langle S' \rangle$ differs at most from $\langle S \rangle$ in the assignment to unmarked gaps.

Basically, we check whether Γ extended with the answer and the antecedents of the answer (assuming they can be added without losing consistency) provides a positive or negative answer to the question.

Notice, that this definition also allows us to deal with answers containing universal quantifiers, since these are also represented as Π-types. For instance, the answer 'Everyone' to the question 'Who walks?' translates into $g : \Pi x : person.walk \cdot x$. Checking whether this is an answer according to the definition amounts to temporarily extending the context with $x : person$ and $g : \Pi x : person.walk \cdot x$. Notice that after that operation, a substitution can be found for the formal representation of the question ($\mathbf{X} : person, P : walk \cdot \mathbf{X}$). The substitution in question is $\langle \mathbf{X} := x, P := g \cdot x \rangle$.

We have now defined conditional answers. We will use the definition as a basis for a full definition of indirect answerhood. A conditional answer can be seen as the most rudimentary type of indirect answer. Indirect answers can be defined as follows on the basis of conditional answer: an indirect answer is an answer which, when added to the context, allows us to derive a (new) conditional answer. According to this definition, conditional answers are also indirect answers. Before we present the formal version of the definition, we first define the notion of contextual equivalence. This notion is needed in the definition of indirect answers to avoid the problem of conversational equivalence.

D. 11 (Contextual equivalence): Segment A is contextually equivalent to segment B in context Γ iff

1. $\Gamma \otimes A \vdash B$; and

2. $\Gamma \otimes B \vdash A$.

D. 12 (Indirect answer): A segment A is an indirect answer for an open segment Q in a context Γ, iff there is an A' such that:

1. $\Gamma \otimes A \vdash_{\Delta} A'$;

2. (consistency) $\Gamma \otimes A$ is consistent;

3. (informativity) It is not the case that $\Gamma \vdash_{\Delta} A'$;

4. There is at least one substitution $\langle S \rangle$ such that

   (a) A' is a conditional answer to Q in the context Γ for substitution S; and

   (b) There is no $\langle S' \rangle$ which differs at most from $\langle S \rangle$ in the assignment to unmarked gaps such that $\Phi(A')$ and $Q\langle S' \rangle$ are contextually equivalent with respect to $\Gamma \otimes A$.

Let us give an abstract example of an indirect answer. Suppose that Γ contains the following introduction: $f : \Pi\langle x : \alpha, y : \beta, z : \gamma\rangle.\delta$. The question at stake is *whether* δ, and the informer provides the answer $a : \alpha$. If we extend the context with the answer, then we can derive $f \cdot a : \Pi\langle y : \beta, z : \gamma\rangle.\delta$, which is a conditional answer with respect to $\Gamma \otimes a : \alpha$ (and could not be derived in Γ). In Groenendijk and Stokhof 1984, the fact that $a : \alpha$ is an indirect answer is accounted for by the fact that in the context extended with this answer the question *whether (if* β *and* γ *then* δ*)* becomes relevant.

The question now arises of whether our definition of indirect answerhood covers all instances of indirect answers that Groenendijk and Stokhof cover with their definition. According to their definition, an indirect answer A leads to a context in which there is at least one question which was not present in the previous context and whose answer X also (partially) answers the original question Q. In our framework, this means that there should be a substitution $\langle S\rangle$ such that (conditions on informativity and consistency are not relevant at this point of the discussion and therefore left implicit):

$\Gamma \otimes A \otimes X \vdash_\Delta Q\langle S\rangle$; or
$\Gamma \otimes A \otimes X \vdash_\Delta \mathit{NOT(Q)}\langle S\rangle$.

In words, the context $\Gamma \otimes A$ extended with the answer X to the new question, should yield a positive or negative answer to the original question Q. But this corresponds to having a substitution $\langle S\rangle$ such that:[9]

$\Gamma \otimes A \vdash_\Delta p : \Pi\langle X\rangle.Q\langle S\rangle$; or
$\Gamma \otimes A \vdash_\Delta p : \Pi\langle X\rangle.\mathit{NOT(Q)}\langle S\rangle$.

The two formulae after $\vdash_\Delta$ are in fact, according to our definitions, conditional answers to Q. Therefore, the answer A fulfills the criteria for being an indirect answer: adding A to the context yields a context in which a conditional answer can be derived.

The next question that may be raised is whether our definition is also infected by the problem of conversational equivalence. Reconsider the situation in which the two atomic propositions ϕ and ψ are independent with respect to some context Γ. Somebody asks *whether* ψ. Let us see whether according to our definition of indirect answers, ϕ is an indirect answer in this situation. At first sight, it does indeed seem to be so, since there seems to be a conditional answer to *whether* ψ which can be derived in $\Gamma \otimes p : \phi$. The type of the answer is $(\phi \to \psi) \to \psi$.[10] This answer meets the condition 4.(a) and thus definition 10: if this answer and a proof for its antecedent $(\phi \to \psi)$ are added to $\Gamma \otimes p : \phi$, then a proof for ψ can be derived. We have, however, not yet taken the condition 4.(b) for indirect answers into account; it says that the substition instance thus obtained of *whether* ψ*)* –i.e., ψ– should not be contextually equivalent to $\Phi((\phi \to \psi) \to \psi) = (\phi \to \psi)$ with respect to $\Gamma \otimes A$. It can, however be easily verified that they are contextually equivalent (for the reader's convience, we have ommited the proof objects):

[9] Compare it with the following equivalence for propositional logic: $\Gamma, p \vdash q \Leftrightarrow \Gamma \vdash p \to q$.

[10] We have $p : \phi \vdash (\lambda q : \phi \to \psi.q \cdot p) : (\phi \to \psi) \to \psi$. This follows (using CTT's abstraction rule, which is closely related to arrow introduction in propositional logic) from $p : \phi, q : \phi \to \psi \vdash q \cdot p : \psi$.

$\Gamma \otimes \phi \otimes \phi \rightarrow \psi \vdash \psi$ and
$\Gamma \otimes \phi \otimes \psi \vdash \phi \rightarrow \psi$.

Thus, we may conclude that the condition 4.(b) on contextual equivalence rules out ϕ as an answer to *whether* ψ and frees us from the problem of conversational equivalence.

Finally, we would like to note that there are responses to questions which may be termed answers, which Groenendijk and Stokhof 1984 do not deal with. Let us give an example. Suppose that it can be derived in Γ that *if* ϕ *then* ψ. Furthermore, the question again is *whether* ψ. Now, the assertion of *not* ϕ seems informative for the questioner: this answer rules out one line of investigation for the questioner, i.e., looking for an answer to ϕ in order to find an answer to ψ. We will call this sort of answer a *preventive* answer, because it is intended to prevent the questioner from searching for an answer in the wrong direction. In our framework, this sort of response or answer can be easily formalized:

D. 13 (preventive answer): Segment A is a preventive answer to open segment Q in Γ iff there is some segment A′ such that:

1. A' answers Q in Γ;
2. A' does not answer Q in $\Gamma \otimes A$.

5.5 A remark on a computational model

In this section, we indicate how to turn our framework into a computational model of answerhood. Our framework is based on a proof system which can be implemented. It suffers, however, from the problem that any proof system suffers from that deals with a logic which is as strong as predicate logic: it is not decidable. This means that we need heuristics in order to make the system run properly (i.e., not go into a non-terminating search process now and again).

The heuristics (for instance, specifying the search depth or order of search) can be seen as being associated with an agent $\mathcal{A}$. In our definitions, we can then replace any occurrence of $\Gamma \vdash_{\Delta} C$ with

Agent $\mathcal{A}$ can compute $\Gamma \vdash_{\Delta} C$.

This means that whether something counts as an answer comes to depend on the processing capabilities of $\mathcal{A}$, and thus the notion of answerhood is relativised with respect to the questioner. In other words, answers which I cannot grasp are no answers to me.

6 Conclusions

CTT is suitable for formalizing the notion of answerhood in such a way that the context-dependence of answers (see section 2 for a description of four sorts of context-dependence) is properly accounted for. The formalization which has been presented is based on the idea that a question presents a gap in the information of the questioner, and that determining whether an answer fills a gap (in a context) comes down to performing a deduction in a proof system.

Our approach is shown to build on two types of theories about answerhood which are well-represented in the literature. First, we show that our approach generalizes those theories that start from the idea that answerhood should be explicated in terms of the possibility to unify a question and its answer (e.g., Katz 1968, Scha 1983). Second, the dynamic and contextual side of answerhood, as reflected in Groenendijk and Stokhof's definition of answerhood, is formalized. The formalization on the basis of CTT allows for a finer-grained analysis when it comes to dealing with indirect answers. The fact that our approach operates on the logical (syntactic) form of answers (as opposed to the operations on possible worlds as employed by Groenendijk and Stokhof 1984) allows to isolate conditional answers from indirect answers. The former are an interesting class of answers which are computionally more feasible than indirect answers. Furthermore, we show how the notion of conversational equivalence can be formalised in our framework.

The framework which we have described is a generalisation of DRT.[11] This means that, for instance, the treatment of anaphora fits nicely into it. Also the treatment of presuppositions as anaphora can be worked out in the framework (see, e.g., Krahmer and Piwek 1997). Here there is a difference with the possible-worlds framework employed by Groenendijk and Stokhof 1984 which is not suited for these purposes.[12]

Finally, a practical advantage of the formalization that has been presented is that the formal definition can be fitted into a computational model, which is applicable in question-answering systems of the future.

References

Ahn, R.: 1994, Communicating contexts: A pragmatic approach to information exchange, in P. Dybjer, B. Nordström, and J. Smith (eds.), *The proceedings of the BRA workshop: Types of Proofs and Programs*, Vol. 996 of *Springer Lecture Notes in Computer Science*, Springer Verlag, Berlin

Ahn, R. and Kolb, H.-P.: 1990, Discourse representation meets constructive mathematics, in L. Kálmán and L. Pólos (eds.), *Papers from the Second Symposium on Logic and Language*, pp 105–124, Akademiai Kiado, Budapest

Barendregt, H.: 1992, Lambda calculi with types, in S. Abramsky, D. Gabbay, and T.

[11] In DRT truth-conditions play an important role. We have not talked about these with respect to CTT. One does, however, want to distinguish true and false answers. To draw this distinction, an answer has to be evaluated with respect to a language external reality. This could, for instance, be done by exploiting the translation between CTT contexts and DRSs (Ahn and Kolb 1990). Alternatively, one could check whether the answer can be derived in the CTT context of some omniscient being Ω (i.e., God, cf. Ahn and Kolb 1990).

[12] However, Groenendijk and Stokhof are currently extending their framework in this direction.

Maibaum (eds.), *Handbook of Logic in Computer Science*, Oxford University Press, Oxford

Bäuerle, R. and Zimmermann, T.: 1991, Fragesätze, in A. von Stechow and D. Wunderlich (eds.), *Semantik/Semantics: An International Handbook of Contemporary Research*, pp 333–348, Walter de Gruyter, Berlin

Cohen, F.: 1929, What is a question?, *The Monist* 39, 350–364

Curry, H. and Feys, R.: 1958, *Combinatory Logic*, Vol. 1, North-Holland, Amsterdam

De Bruijn, N.: 1980, A survey of the project automath, in J. Seldin and J. Hindley (eds.), *To H.B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalisms*, pp 589–606, Academic Press

Groenendijk, J. and Stokhof, M.: 1984, On the semantics of questions and the pragmatics of answers, in F. Landman and F. Veltman (eds.), *Varieties of Formal Semantics*, pp 143–170, Foris, Dordrecht

Groenendijk, J. and Stokhof, M.: 1997, Questions, in J. van Benthem and A. ter Meulen (eds.), *Handbook of Logic and Language*, pp 1057–1124, North-Holland, Amsterdam

Hamblin, C.: 1971, Mathematical models of dialogue, *Theoria* 37, 130–155

Hintikka, J.: 1974, Questions about questions, in M. Munitz and P. Unger (eds.), *Semantics and Philosophy*, pp 103–158, New York University Press, New York

Jespersen, O.: 1933, *Essentials of English Grammar*, George Allen and Unwin Ltd., London

Kamp, H.: 1981, A theory of truth and semantic interpretation, in J. Groenendijk, T. Janssen, and M. Stokhof (eds.), *Formal Methods in Study of Language*, Mathematical Centre Tracts 136, pp 277–322, Foris, Amsterdam

Kamp, H. and Reyle, U.: 1993, *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Kluwer Academic Publishers, Dordrecht

Katz, J.: 1968, The logic of questions, in B. van Rootselaar and J. Staal (eds.), *Logic, Methodology and Philosophy of Science III: Proceedings of the Third International Congress for Logic, Methodology and Philosophy of Science*, pp 463–493, North-Holland, Amsterdam

Krahmer, E. and Piwek, P.: 1997, Presupposition projection as proof construction, in H. Bunt, L. Kievit, R. Muskens, and M. Verlinden (eds.), *Proceedings of the Second International Workshop on Computational Semantics*, pp 122–134, Tilburg

Martin-Löf, P.: 1984, *Intuitionistic Type Theory*, Bibliopolis, Naples

Piwek, P.: 1997, *What is an answer? That is the question*, forthcoming

Prüst, H., Scha, R., and Van den Berg, M.: 1994, Discourse grammar and verb phrase anaphora, *Linguistics and Philosophy* 17(3), 261–327

Ranta, A.: 1990, Anaphora in game theoretical semantics and in intuitionistic type theory, in L. Haaparanta, M. Kusch, and I. Niiniluoto (eds.), *Language, Knowledge and Intentionality: Perspectives on the Philosophy of Jaakko Hintikka*, Acta Philosophica Fennica, pp 265–274, Helsinki

Scha, R.: 1983, *Logical Foundations for Question Answering*, Ph.D. thesis, University of Groningen, Groningen

Stalnaker, R.: 1974, Pragmatic presupposition, in M. Munitz and P. Unger (eds.), *Semantics and Philosophy*, pp 197–213, New York University Press, New York

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

Updating and revising the agents mental state in dialogues

Paulo Quaresma and José Gabriel Lopes

{pq,gpl}@di.fct.unl.pt
Departamento de Informática,
Faculdade de Ciências e Tecnologia,
Universidade Nova de Lisboa,
2825 Monte da Caparica
Portugal

Abstract

In this paper we present an extended logic programming framework that allows to model dialogues between different kinds of agents. Namely, it will be shown how this framework is able to handle dialogues between agents with different levels of sincerity, cooperativeness, credulity, and activity.

In this framework an agent/computational system is modeled by a set of extended logic programming rules representing its mental state. These rules describe the agent behavior, attitudes (believes, intentions, and objectives), world knowledge, and temporal and reasoning procedures. The complete mental state is defined by the well founded model of the extended logic program that models the agent.

Using this modeling process an agent is able to participate in dialogues, updating and revising its mental state after each sentence. The revision process includes the capability to remove contradictions in the agent mental state.

1 Introduction

In order to participate in dialogues, an agent/computational system needs the capability to model its mental state. Namely, it is necessary to represent the agent attitudes (believes, intentions, and objectives), world knowledge and temporal, reasoning and behavior rules. In this paper, we propose a logic programming framework that allows the representation of agent models and the definition of update and revise procedures. In a dialogue, these procedures are executed after each event (sentence) and they update the agent model using the information associated with the different speech acts.

Agent models are defined as logic programs extended with explicit negation and the semantics of the programs is given by the well founded semantics of logic programs with explicit negation (from Pereira et al. Alferes and Pereira 1996; Alferes et al. 1995; Alferes

1993). The well founded semantics has a complete and sound top-down proof procedure with polynomial complexity and there is an implemented prototype (Damásio et al. 1994) which allows us to obtain experimental results.

At each time instant, the agent mental state is given by the well founded model of the logic program that models the agent. In a dialogue, after each sentence, it is necessary to update the agent model with the new information. This process is done through the update of the logic program with the facts that describe the events: identification of the time and speech acts associated with each event.

However, the update process may create a contradictory mental state. For instance, it is possible that an event initiates a belief that is contradictory with some previous believes. In these situations, it is necessary to revise the agent mental state, terminating the attitudes that supported the contradiction.

The updated and, eventually, revised agent mental state may be used as the input of a planning procedure that tries to satisfy the agent objectives.

The proposed framework has some advantages over many classical dialogue systems: Litman and Allen (Litman 1985; Litman and Allen 1987; Allen et al. 1991), Carberry (Carberry 1985; Carberry 1988), Pollack (Pollack 1986; Pollack 1990). In fact, it supports the recognition of attitudes using a formal framework with a specific semantic. Moreover, it allows the representation of several kinds of users and it supports the existence of contradictory states, eliminating the contradiction when necessary. These characteristics allow this framework to handle a wider range of dialogues, dealing with error situations and non-well behaved agents.

In the next section, the logic programming framework is briefly described. In section 3 we present the agent modelling process, with a special focus on the capability to model agents with different levels of sincerity, cooperativeness, credulity, and activity. The procedures to update and revise the agents mental state after each event are described in section 4 and 5. The planning recognition process is described in section 6. Finally, in section 7 some conclusions and open problems are pointed out.

2 Logic programming framework

Logic programs extended with explicit negation are finite set of rules of the form

- $H \leftarrow B_1, \ldots, B_n, \text{not } C_1, \ldots, \text{not } C_m \quad (m \geq 0, n \geq 0)$

where H, B_1, ..., B_n, C_1,..., C_m are objective literals. An objective literal is an atom A or its explicit negation $\neg A$; *not* stands for negation by default; *not* L is a default literal. Literals are objective or default and $\neg\neg L \equiv L$.

The set of all ground objective literals of a program P designates the extended Herbrand base of P and it is represented by $H(P)$. An interpretation I of an extended program P is represented by $T \cup \mathit{not}\ F$, where T and F are disjoint subsets of $H(P)$. Objective literals of T are true in I; objective literals of F are false by default in I; objective literals of $H(P) - I$ are undefined in I. Moreover, if $\neg L \in T$ then $L \in F$.

An interpretation I of an extended logic program P is a partial stable model of P iff $\Phi_P(I) = I$ (see Alferes and Pereira 1996 for the definition of the Φ operator).

The well founded model of the program P is the F-least partial stable model of P. The well founded semantics of P is determined by the set of all partial stable models of P.

Pereira et al. (Alferes and Pereira 1996; Alferes et al. 1995) showed that every non-contradictory program has a well founded model and they also presented a complete and sound top-down proof procedure for several classes of programs.

In their work, Pereira et al., proposed a revision process that restores consistency for contradictory programs, taking back assumptions of the truth value of negative literals. As it will be described in section 4, we also use this approach in order to revise the agents mental state.

2.1 Events

The agent modeling process must be able to deal with time and events. In fact, it is very important that agents have the capability to reason about their mental state at a given time point. They should also be able to change their mental state as a consequence of some external or internal events.

As a time formalism we propose a variation of the Event Calculus (Shanahan 1989; Eshghi 1988; Missiaen 1991) that allows events to have an identification and a duration. As a consequence events may occur simultaneously.

The predicate *holds_at* defining the properties that are true at a specific time is:

$$
\begin{array}{lrcl}
(1) & \mathit{holds\_at}(P,T) & \leftarrow & \mathit{happens}(E,T_i,T_f), \\
 & & & \mathit{initiates}(E,T_P,P), \\
 & & & T_P < T, \\
 & & & \mathit{persists}(T_P,P,T). \\
(2) & \mathit{persists}(T_P,P,T) & \leftarrow & \mathit{not\ clipped}(T_P,P,T). \\
(3) & \mathit{clipped}(T_P,P,T) & \leftarrow & \mathit{happens}(C,T_{ci},T_{cf}), \\
 & & & \mathit{terminates}(C,T_C,P), \\
 & & & \mathit{not\ out}(T_C,T_P,T). \\
(4) & \mathit{out}(T_C,T_P,T) & \leftarrow & T \leq T_C. \\
(5) & \mathit{out}(T_C,T_P,T) & \leftarrow & T_C < T_P.
\end{array}
$$

The predicate $\mathit{happens}(E,T_i,T_f)$ means that the event E occurred between T_i and T_f; $\mathit{initiates}(E,T,P)$ means that the event E initiates P at time T; $\mathit{terminates}(E,T,P)$ means that event E terminates P at time T; $\mathit{persists}(T_i, P, T)$ means that P persists since T_i until T (at least); $\mathit{succeeds}(E,T_i)$ means that the event E may occur at time T_i (its pre-conditions are satisfied).

Note that a property P is true at a time T ($holds\_at(P,T)$), if there is a previous event that initiates P and if P persists until T. P persists until T if it can not be proved by default the existence of another event that terminates P before the time T.

We need additional rules for the relation between not holding a property and holding its negation and we also need to define the relation between the two kinds of negation:

$$\neg holds\_at(P,T) \leftarrow holds\_at(\neg P,T). \quad (6)$$

$$\neg holds\_at(P,T) \leftarrow not\ holds\_at(P,T). \quad (7)$$

The predicates need to be related by some integrity rules:

1. Events can not initiate and terminate a property at the same time:

   $$\leftarrow initiates(E,T,P),\ terminates(E,T,P). \quad (8)$$

2. Events can not initiate/terminate a property and its negation:

   $$\leftarrow initiates(E,T,P), initiates(E,T,\neg P). \quad (9)$$

   $$\leftarrow terminates(E,T,P), terminates(E,T,\neg P). \quad (10)$$

3. Events can not be associated to different time intervals:

   $$\begin{array}{rl} \leftarrow & happens(E,T_{1i},T_{if}), \\ & happens(E,T_{2i},T_{2f}), \\ & T_{1i} = T_{2i}, \\ & not(T_{if} = T_{2f}). \end{array} \quad (11)$$

4. Events can not have a negative duration:

   $$\leftarrow happens(E,T_i,T_f),\ not(T_i \leq T_f). \quad (12)$$

5. Events must have an associated action:

   $$\begin{array}{rl} \leftarrow & happens(E,T_i,T_f), \\ & not(act(E,A)). \end{array} \quad (13)$$

6. Properties must be initiated by some event:

   $$\begin{array}{rcl} & \leftarrow & holds\_at(P,T), \\ & & not(ev\_gen(P,T)). \\ ev\_gen(P,T) & \leftarrow & happens(E,T_i,T_f), \\ & & initiates(E,T_p,P), \\ & & T_i \leq T_p \leq T, \\ & & persists(T_p,P,T). \end{array} \quad (14)$$

7. Events can not occur if the pre-conditions are not satisfied:

   $$\leftarrow happens(E,T_i,T_f),\ not\ succeeds(E,T_i). \quad (15)$$

3 Agents mental state

In our proposal, agents are modeled by the well founded model of an extended logic program with the following structure:

1. Rationality rules (RR). These rules describe the relation between the different attitudes (believes, intentions, and objectives).
2. Behavior rules (BR). These rules define the agent activity, cooperativeness, credulity, and sincerity.
3. Actions description (Ac). These rules describe the actions that may be executed by the agent. In the domain of dialogues, these rules describe the speech acts, their pre-conditions and effects.
4. A temporal formalism (T). These are the rules presented in the previous section.
5. World knowledge (WK).These rules describe the agent world knowledge: entities, taxonomies, ...

In the next subsections we will analyze the first two structures: rationality rules and behavior rules.

3.1 Rationality rules

These rules define relations between the agents attitudes: believes (bel), objectives (ach), and intentions (int).

The main relations are (for related work see Bratman 1990; Cohen and Levesque 1990a; Cohen and Levesque 1990b; Perrault 1990):

- Integrity

(16) $$\perp \leftarrow holds\_at(bel(A,P),T),\ holds\_at(bel(A,\neg P),T).$$

(17) $$\perp \leftarrow holds\_at(ach(A,P),T),\ holds\_at(ach(A,\neg P),T).$$

- Consistency

(18) $$\neg holds\_at(bel(A,\neg P),T) \leftarrow holds\_at(bel(A,P),T).$$

(19) $$\neg holds\_at(ach(A,\neg P),T) \leftarrow holds\_at(ach(A,P),T).$$

- Introspection

(20) $$\perp \leftarrow holds\_at(bel(A,P),T), holds\_at(bel(A,\neg bel(A,P)),T).$$

(21) $$\perp \leftarrow \neg holds\_at(bel(A,P),T), holds\_at(bel(A,bel(A,P)),T).$$

- Necessity

(22) $$holds\_at(bel(A,P),T) \leftarrow holds\_at(P,T).$$

3.2 Behavior rules

These rules allow the definition of the agent behavior. As behavior properties we have considered the credulity, sincerity, activity, and cooperativeness.

The behavior properties are represented by agent believes (about himself and about the other agents). For instance, an agent a believes at a time t that the agent b is sincere, credulous, cooperative and reactive.

$$\begin{array}{l} holds\_at(bel(a, sincere(b)), t). \\ holds\_at(bel(a, credulous(b)), t). \\ holds\_at(bel(a, cooperative(b)), t). \\ holds\_at(bel(a, reactive(b)), t). \end{array}$$

Note that this approach allows an agent to have different behaviors depending of the time instant. Moreover, an agent may change his believes about the others behavior. However, in this paper we will not discuss the events that may contribute to these changes.

3.2.1 Credulity

Credulity defines how an agent accepts information from other agents. The main process defines how believes are transferred:

$$(23) \qquad holds\_at(bel(H, P), T) \leftarrow \begin{array}{l} holds\_at(bel(H, bel(S, P)), T), \\ holds\_at(bel(H, credulous(H)), T). \end{array}$$

This rule defines that an agent believes in a proposition if he believes that another agent believes in it and if he is credulous (at that time). This rule can be changed for a more sceptical agent adding more pre-conditions (check if the belief is not contradictory with some previous belief).

The credulity property has also consequences over the description of the speech acts. In fact, the effect of a speech act depends on the model that the hearer has of the speaker.

In this paper, we will show only the effect of the $inform$ speech act:

$$(24) \qquad initiates(E, T_f, bel(H, bel(S, P))) \leftarrow \begin{array}{l} act(E, inform(S, H, P)), \\ happens(E, T_i, T_f), \\ holds\_at(bel(H, sincere(S)), T). \end{array}$$

3.2.2 Sincerity

Sincerity defines what is the relation between what agents speak and what they believe.

The effect of a $inform$ speech act for a sincere agent is:

$$(25) \quad initiates(E, T_f, bel(S, bel(H, bel(S, P)))) \leftarrow act(E, inform(S, H, P)),$$

$$\begin{aligned} & happens(E, T_i, T_f), \\ & holds\_at(bel(S, sincere(S)), T), \\ & holds\_at(bel(S, P), T). \end{aligned}$$

For a non sincere speaker, the effect will be (note that the speaker informs the incorrect truth value of P):

$$(26) \qquad initiates(E, T_f, bel(S, bel(H, bel(S, \neg P)))) \leftarrow \begin{aligned}[t] & act(E, inform(S, H, \neg P)), \\ & happens(E, T_i, T_f), \\ & holds\_at(bel(S, sincere(S)), T), \\ & holds\_at(bel(S, P), T). \end{aligned}$$

3.2.3 Cooperativeness

This property defines how intentions and objectives are transferred between agents.

For a cooperative agent:

$$(27) \qquad holds\_a(int(H, A), T) \leftarrow \begin{aligned}[t] & holds\_at(bel(H, int(S, A)), T), \\ & holds\_at(bel(H, cooperative(H)), T). \end{aligned}$$

$$(28) \qquad holds\_a(ach(H, P), T) \leftarrow \begin{aligned}[t] & holds\_at(bel(H, ach(S, P)), T), \\ & holds\_at(bel(H, cooperative(H)), T). \end{aligned}$$

For a non cooperative agent:

$$(29) \qquad holds\_a(\neg int(H, A), T) \leftarrow \begin{aligned}[t] & holds\_at(bel(H, int(S, A)), T), \\ & holds\_at(bel(H, non\_cooperative(H)), T). \end{aligned}$$

$$(30) \qquad holds\_a(\neg ach(H, P), T) \leftarrow \begin{aligned}[t] & holds\_at(bel(H, ach(S, P)), T), \\ & holds\_at(bel(H, non\_cooperative(H)), T). \end{aligned}$$

3.2.4 Activity

An agent may be pro-active or reactive. A pro-active agent has some objectives that were not transferred from other agents: the agent has his own motivations. On the contrary, a reactive agent acts only as a response to other agents acts.

A pro-active agent as the following property (he has his own objectives):

(31) $$holds\_at(bel(A, proactive(A)), T) \leftarrow holds\_at(ach(A, P), T), \; not\ holds\_at(bel(A, ach(B, P)), \; not(A = B).$$

4 Updating an agent mental state

The agent mental state, as it was defined in the previous sections, must be updated after each event.

This process is defined in the following way:

Definition 9: Let m be the agent model, : $m = < RR, Ac, T, BR, WK >$, where RR are the rationality rules, Ac are the rules defining the domain actions, T are the temporal axioms, BR are the behavior rules and WK are the works knowledge rules.

The update function $update : M \times E^n \longrightarrow M$, is defined such as:

1. $update(m, e_1 \times ... \times e_n) = < RR_1, Ac_1, T_1, BR_1, WK_1 >$
2. $RR_1 = RR$
3. $Ac_1 = Ac$
4. $T1 = T$
5. $BR_1 = BR$
6. $WK_1 = WK \cup \{act(e_1, a_1), happens(e_1, t_1, t'_1)..., act(e_{n,} a_n), happens(e_n, t_n, t'_n)\}$, the world description is updated with the new events $e_1, ..., e_n$.

The new agent attitudes are the properties at ($bel/2$, $ach/2$, and $int/2$) that hold in the new model:

$$holds\_at(at, t) \in WFM(RR \cup Ac \cup T \cup BR \cup WK_1)$$

The update process may initiate some attitudes which are inconsistent with the previous mental state. In this situation, the model must be revised, and some attitudes should be terminated. In the next section this process will be described.

Using this update process it is possible to handle situations where the agent believes in one property at a given time point, and then he changes his belief. Note that the previous belief is not lost (the agent knows at which time interval it hold).

As an example, agent a believes at time t_1, that Kathy is at the hospital:

$$holds\_at(bel(a, at(hospital, kathy)), t_1).$$

At a greater time point, $t_2 > t_1$, he is informed that she is at home:

$$happens(e_1, t_2, t_2).$$
$$act(e_1, inform(b, a, at(home, kathy))).$$

If agent a is credulous, he will adopt the new information (using the speech act $inform$ presented previously):

1. $$holds\_at(bel(a, at(hospital, kathy)), t_2).$$

2. $$holds\_at(bel(a, at(home, kathy)), t_2).$$

However, there should be an integrity constraint stating that is contradictory to believe that an agent may be at two different places at the same time:

$$\begin{aligned}\perp \;\; \leftarrow & \; holds\_at(bel(A, at(B, L_1)), T), \\ & holds\_at(bel(A, at(B, L_2)), T), \\ & L_1 \neq L_2.\end{aligned}$$

In this situation, the model must be revised and one of the non-contradictory solution must be chosen (see next section).

5 Revising Mental States

As it was shown, the update process may introduce contradiction in the agents mental state. In fact, the new events ($happens(e_i, t, t'), act(e_i, a_i)$, with $1 \leq i \leq n$), may introduce contradiction due to two different causes:

1. Contradiction caused by the new facts;

2. Contradiction caused by the effects of the new facts.

The first type of contradiction is caused by the violation of integrity constraints relating the description of facts (happens and act) and is analyzed at subsection 5.1. The second cause of contradiction is associated with the effects of the new events (see section 5.2).

5.1 Contradictory facts

Contradiction may be caused by the description of the new events. As an example, suppose the open window action (very simplified version):

$$\begin{array}{lrcl} (32) & enabled(E, T_i) & \leftarrow & act(E, open\_window), \\ & & & holds\_at(closed\_window, T_i). \\ (33) & initiates(E, T_f, opened\_window) & \leftarrow & happens(E, T_i, T_f), \\ & & & act(E, open\_window), \\ & & & holds\_at(closed\_window, T_i). \end{array}$$

Suppose the window is open at time t_0:

$$\begin{array}{ll} (34) & happens(e_0, t_0, t_0). \\ (35) & act(e_0, start). \\ (36) & initiates(e_0, t_0, opened\_window). \end{array}$$

The agent recognized the following event:

$$\begin{array}{ll} (37) & happens(e, t, t'). \\ (38) & act(e, open\_window), \\ (39) & t_0 \leq t \leq t'. \end{array}$$

In these conditions, the model is contradictory because the integrity constraint 15 is violated:

$$\perp \ \leftarrow happens(E, T_i, T_f), not\ enabled(E, T_i).$$

In fact, it is not possible to infer $enabled(e, t)$, because the following property can not be inferred:

$$holds\_at(janela\_fechada, t).$$

Contradiction can be removed using two approaches:

1. Abducting one action that allows the satisfaction of some desired properties (for instance, a previous example that had closed the window);

2. Assuming that there was an incorrect event recognition.

The first approach is already supported by the proposed framework through the use of the rules that allow the abduction of events ($happens/3$ and $act/2$). These rules allow to avoid contradiction whenever is possible to abduce actions that create the desired conditions.

However this process does not guarantee that the model is always non-contradictory. In this case it is necessary to use the contradiction removal process described by Alferes and Pereira 1996. The revisable predicates are:

$$rev = \{not\ happens(E, T_i, T_f), not\ act(E, A)\}$$

This set of revisables mean that is possible to revise the existence of events and its associated actions, whenever necessary. If the new events can not be incorporated in the agent model, then its existence should be revised (the model should always be non-contradictory).

The revision process allows the definition of the preferred non-contradictory solutions (Damásio et al. 1994). We can define the revision process to revise the newest/oldest events that support contradiction.

5.2 Contradictory Mental State

Contradiction may also be cause by the effects of the new events (these effects may violate some integrity constraints).

This kind of contradiction can be detected through the calculus of the contradiction support set of an extended logic program (Alferes and Pereira 1996). Contradictions caused by the effects of events are associated with integrity constraint rules of the following form:

$$\perp\ \leftarrow holds\_at(P_1, T), holds\_at(P_2, T).$$

In this situation, the revising process allows the definition of preference rules over the properties (preferring a non-contradictory solution where some properties hold). Suppose that an event initiated property P_2, and property P_1 is also valid; a possible approach could be to revise P_1 (or P_2) revising the assumption that it hasn't terminated.

As an example, suppose the situation of the previous section where agent a believes Kathy is at the hospital and he is informed that she is at home:

$$\begin{array}{l} holds\_at(bel(a, em(hospital, kathy)), t_2). \\ holds\_at(bel(a, em(home, kathy)), t_2). \end{array}$$

We have the integrity constraint:

$$(40) \qquad \perp\ \leftarrow \quad \begin{array}{l} holds\_at(bel(X, at(L_1, Y)), T), \\ holds\_at(bel(X, at(L_2, Y), T), \\ not(L_1 = L_2). \end{array}$$

The revision process obtains the non-contradictory solutions and it also obtains the preferred solution (accordingly with a pre-defined order between properties):

1. Terminate the believe that Kathy is at the hospital;

2. Terminate the believe that Kathy is at home.

6 Plan Recognition

The update and revise processes presented in the previous sections allow the definition of the agent mental state, after each event. Using this model it is possible to try to recognize the other agents plans in order to participate actively in the interaction process.

An agent plan, at a given time, is the set of intentions about the actions he wants to be realized:

Definition 10: ¿From the agent a point of view, $P_a(b,t)$ is the agent's b plan at time t, and is defined by:

$$P_a(b,t) = \{int(b,X) : holds\_at(bel(a, int(b,X)), t) \in WFM(M_a)\}$$

where M_a is the model of a.

After each event, an agent tries to recognize the other agents plans and he uses them in order to plan his own actions.

An agent plan is given by $P_a(a,t)$, representing the actions he wants to be performed:

$$P_a(a,t) = \{int(a,X) : holds\_at(bel(a, int(a,X)), t) \in WFM(M_a)\}$$

Plan recognition is, in this framework, the inference of the believes about intentions. This plan recognition strategy may be seen as a mixture of the classical-STRIPS approaches (Fikes and Nilsson 1971; Litman 1985) and the mental states approach of Pollack (Pollack 1990). Actions are described in terms of their pre-conditions and effects (as in classical planning), but the recognition process is based on a attitude theory defining the agents mental states and their relation with the speech acts.

The next step is to generate an agent plan. This process is done using an abductive planning strategy that abduces the actions needed to create the desired states. However this process is not described in this paper (see Quaresma and Lopes 1995; Quaresma 1997).

7 Conclusions

We have proposed an agent modeling process with the following characteristics:

1. It was defined over a logic programming framework with a specific semantic (well founded semantics of extended logic programs);

2. It has a complete and sound top-down proof procedure;

3. It allows the definition of reasoning and behavior rules. These rules allow the modeling of non-well behaved agents;

4. It has an update and revise procedure defined for any event that may occur;

5. It may be the base of a planning process that allows the participation of agents in dialogues.

This framework has some advantages over previous systems because it is formal, it may be implemented (in fact we have a working prototype), and it supports a wide range of dialogue situations.

However, there are many problems to be deal as future work.

First, and as it was pointed out in the previous section, we have not analyzed the integration of the modeling process with the planning process and the natural language generation phase. Moreover, it was not discussed the problem of the recognition of speech acts from natural language sentences. These tasks are pre-conditions for the construction of a robust natural language processing system.

As future work we also intend to integrate this agent modeling framework in a more general architecture allowing a complete representation of dialogues. Namely, the architecture should be able to deal with cycles of conversation and clarification dialogues (Lopes 1991; Quaresma and Lopes 1992).

References

Alferes, J. J.: 1993, *Semantics of Logic Programs with Explicit Negation*, Ph.D. thesis, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

Alferes, J. J., Damásio, C., and Pereira, L. M.: 1995, A logic programming system for nonmonotonic reasoning, *Journal of Automated Reasoning* 14, 93–147

Alferes, J. J. and Pereira, L. M.: 1996, *Reasoning with Logic Programming*, Vol. 1111 of *Lecture Notes in Artificial Intelligence*, Springer

Allen, J., Kautz, H., Pelavin, R., and Tenenberg, J.: 1991, *Reasoning about Plans*, Morgan Kaufman Publishers, Inc.

Bratman, M.: 1990, *What is Intention?, in Intentions in Communication*, MIT

Carberry, S.: 1985, *Pragmatic Modeling in Information System Interfaces*, Ph.D. thesis, University of Delaware

Carberry, S.: 1988, Modelling the user's plans and goals, *Computational Linguistics* 14(3), 23–37

Cohen, P. and Levesque, H.: 1990a, Intention is choice with commitment, *Artificial Intelligence* 42(3)

Cohen, P. and Levesque, H.: 1990b, *Persistence, Intention, and Commitment, in Intentions in Communication*, pp 33–70, MIT

Damásio, C., Nejdl, W., and Pereira, L. M.: 1994, Revise: An extended logic progamming system for revising knowledge bases, in M. Kaufmann (ed.), *KR'94*

Eshghi, K.: 1988, Abductive planning with event calculus, in *Proceedings of the International Conference on Logic Programming*

Fikes, R. E. and Nilsson, N. J.: 1971, Strips: A new approach to the application of theorem proving to problem solving, *Artificial Intellligence* (2), 189–208

Litman, D. and Allen, J.: 1987, A plan recognition model for subdialogues in conversations, *Cognitive Science* (11), 163–200

Litman, D. J.: 1985, *Plan Recognition and Discourse Analysis: An Integrated Approach for Understanding Dialogues*, Ph.D. thesis, Dep. of Computer Science, University of Rochester

Lopes, J. G.: 1991, Architecture for intentional participation of natural language interfaces in conversations, in C. Brown and G. Koch (eds.), *Natural Language Understanding and Logic Programming III*, North Holland

Missiaen, L.: 1991, *Localized Abductive Planning with the Event Calculus*, Ph.D. thesis, Univ. Leuven

Perrault, R.: 1990, *An Application of Default Logic to Speech Act Theory, in Intentions in Communication*, Chapt. 9, pp 161–186, MIT

Pollack, M.: 1990, *Plans as Complex Mental Attitudes, in Intentions in Communication*, Chapt. 5, pp 77–104, MIT

Pollack, M. E.: 1986, *Inferring Domain Plans in Question-Answering*, Ph.D. thesis, Dep. of Computer and Information Science, University of Pennsylvania

Quaresma, P.: 1997, *Inferência de Atitudes em Diálogos*, Ph.D. thesis, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, In Portuguese. To appear

Quaresma, P. and Lopes, J. G.: 1992, A two-headed architecture for intelligent multimedia man-machine interaction, in *B. de Boulay and V. Sgurev (eds). Artificial Intelligence V - methodology, systems, applications*, North Holland

Quaresma, P. and Lopes, J. G.: 1995, Unified logic programming approach to the abduction of plans and intentions in information-seeking dialogues, *Journal of Logic Programming* (54)

Shanahan, M. P.: 1989, Prediction is deduction but explanation is abduction, in *Proceedings of the IJCAI*

On Tops and Bottoms: Agents' Coordination of Syntax Production in Dialogue*

Hannes Rieser

Faculty of Linguistics and Literary Studies and
SFB 360 "Situierte Künstliche Kommunikatoren"
(Special Research Unit "Situated Artificial Communicators")
Bielefeld University

1 Levels of Agents' Coordination in Dialogue

The investigation of task–oriented dialogue has been beneficial for the investigation of grounding, mutuality with respect to mental states and the rise of conventions in interacting groups of agents. The question asked in this kind of research was, roughly: "How can agents manage to synchronise their verbal and non–verbal interactions in dialogue, how do they initiate and maintain synchronisation?" In an attempt to answer this question, different answers were given, depending on the special aims of the research involved.

H. Clark and Wilkes–Gibbs (1986/1990), e.g. working with the well known tangram figure experiments, found out that referringly used NPs are introduced into discourse according to a *presentation–acceptance* cycle. This implies that one agent suggests some NP to another agent in order to refer to some tangram figure. It is then negotiated whether this NP is to be used in the sequel. In order to indicate that NPs are not set once and for all but open to further consideration, they are phonologically marked. The marking is used to indicate the suggesting agent's estimation concerning the applicability of the NP under consideration.

In a more recent study, Brennan and Clark (1996) maintained that agents establish a *conceptual pact* about how they conceptualise an object, which, of course, determines lexicalisation.

In two studies based on the maze task, Garrod and Anderson (1987) and Garrod and

* This paper is based on a talk I gave at the MunDial 97, University of Munich, Germany, March 10–12, 1997. Thanks go to Anton Benz, Gerhard Jäger and other participants of the MunDial, who commented upon the things I said there. Most of the topics discussed in the paper are treated in a Bielefeld project on the syntax of spoken discourse (see Gibbon et al. 1995 for further information). Here I am indebted to arguments by Walther Kindt, Susanne Kronenberg, Franz Kummert and Kristina Skuplik. I also got technical help from various sides: The pictures were provided by a project on the dynamics of concepts. Here I have to thank Martin Hoffhenke, Björn Knafla, Bernhard Jung and Ipke Wachsmuth, cf. Wachsmuth and Jung 1996 for details on this project. And last not least I am grateful to Clemens Meier, who did the LaTeX'ing.

Doherty (1994) , it was shown, that agents coordinate on the description of local positions of objects. In addition, the 1994 study revealed that coordination spreads fairly quickly within groups of agents and may thus be considered a convention in the Lewisian (1969) sense.

Coordination is also relevant concerning mental states like belief and knowledge: Heydrich and Rieser (1995) , studying directives and answers by agents in task–oriented dialogue, arrived at the conclusion that agents in discourse point out what they consider to be public information and what must hence go into the common ground.

Using eye–tracking technology, Pomplun, Rieser, Ritter and Velichkovsky (1997) discovered that agents even coordinate their foci of attention, mainly by controlling each others direction of gaze.

These and other empirical studies led to the assumption of various general principles concerning agents' behaviour in dialogue: Clark and Wilkes–Gibbs, e.g. maintain that two principles are active in dialogue, the principle of mutual responsibility and the principle of least collaborative effort. Clark and Brennan postulate a "grounding principle concerning reference", which amounts to the following:

> "When speakers present a reference, they do so provisionally, and they then work with their addressees to establish that it has been understood. When speakers first refer to an object as the loafer, they are proposing to their addressees that it be conceptualised as a loafer. The addressee can ratify the proposal ("okay"), modify it ("you mean the man's shoe?"), or solicit another proposal ("which one?") in the process of grounding that reference"
> (Brennan and Clark 1996, p. 1484).

On their way to capture semantic coordination among agents in dialogue, Garrod and Anderson (1987) established an "output–input–coordination principle". It entails that the current speaker in formulating his utterance will try to match the lexical, semantic and pragmatic decisions used in the interpretation of the previous speakers' last utterance as closely as possible. "Bilateral conformity to the principle quickly produces convergence on a common description schema of the kind observed in the maze game dialogues" (Garrod and Doherty 1994, p. 185). Heydrich and Rieser (1995) based their turn–exchange–model on a "principle of Lewisian coordination": An agent's belief in public information concerning the situation brought about by his last directive is a sufficient condition for the production of his next turn.

In the research reported, semantic and pragmatic matters have been focussed on, although indirectly syntactical form also played some role. But so far, there has been no systematic work done on coordination of morphology and syntax in dialogue. In this paper I will concentrate on matters of syntax, but at least some remarks concerning morphology and the syntax–morphology interface should be made. First a cautionary remark: Agents in dialogue show a language behaviour far more creative than usually assumed. This can be seen from their coining new words related to the task at hand. These *ad–hoc* coinings are then subject to the various principles referred to above, especially to the Clark/Brennan "grounding principle". This does not imply, however, that coordination of morphological forms must needs take place. Speakers may stick to their own morphological options, but if they do so, they still have to coordinate their choice of form with a suitable semantics, perhaps along the lines of the Garrod/Anderson "output–input–coordination principle". Roughly the morphological processes involved either follow the structural rules of the natural language in question or they follow some sort of new rule which is perhaps restricted to spoken language. I will provide examples to make clear the issue involved: In the experimental setting described below, agents have to construct a toy airplane out of a set of wooden materials consisting of bars with holes in them, bolts, nuts, and screw–threaded wooden cubes. At some point of the construction procedure this involves that they must name the parts they need. That is where morphological matters enter. Frequently we have the integration of formal aspects (shape) and functional aspects (role in aggregate building up the airplane) within one compound. The examples in (1) are consonant with morphological composition in German, those under (2) seem to be doubtful:

(1)	Siebener	sevener
	Siebenerstange	sevener–bar
	Siebenerverlängerung	sevener–extension
	Siebenerlangstück	sevener–longpiece
	Siebenerflügel	sevener–wing
	Siebenertragfläche	sevener–wing
(2)	Baufixgelbschraube	Baufix–yellow–screw
	Siebenerlöcherndingern	sevener–holes$_{\text{[dative]}}$ things$_{\text{[dative]}}$

To round off the picture, I mention, thereby anticipating the introduction of phenomena playing a central role in the discussion of syntactic matters, that we also have self–repairs concerning morphological processes as in (3) and (4) as well as cooperative production of compounds as in (5). Below, *I* and *C* stand for *Instructor* and *Constructor*, respectively. The significance of these role indications will become clear in Sect. 2.1.

(3) Siebenerlöchern–äh–plättchen
sevener–holes uhm bars

(4) I: Eine Lorenz Baufixgelbschraube.
A Lorenz Baufix–yellow–screw.
C: [laughs] Aha.
I: Oder eine gelbe Lorenz Baufixschraube.
Or a yellow Lorenz Baufix–screw.

(5) C: Ich habe eine rote Sechskantschraube in der Hand
I hold a red six–edge–screw in my hand
und eine Drei.
and a three.
I: Ja, genau.
Right, exactly so.
I: Löchrige Schiene, so.
with–holes bar kind of.
C: Dreilöchrige Schiene.
with–three–holes bar.

Observe that in (5), the final morphological form is *dreilöchrige/with–three–holes*, where one part, *drei/three*, is contributed by the Constructor, and the other *löchrige/with–holes* is provided by the Instructor, who also does the final compounding. Roughly, we have the Clark–Gibsian "presentation–acceptance cycle" here but with respect to compound–formation.

2 How Agents Coordinate: Joint Turn–Production

The ideas and observations presented here are based on a wide range of empirical data gotten by means of a specific experimental setting described below.

2.1 Experimental Setting and Material Used

The experimental setting we have consists of two agents, an Instructor (I) and a Constructor (C) separated by a screen. The task involved is to build up a toy airplane of type "Baufix" * as shown in Fig. 1.

I has built up the airplane according to a plan. C has all "Baufix"–parts needed for construction. I issues the relevant directives. C follows these directives. The director of the experiment starts and closes sessions and acts as a kind of arbiter.

Using this setting we collected 27 classification dialogues showing the agents' classification procedures concerning the parts of the construction kit, 22 construction dialogues, speech recordings separated for I and C and videos of construction activities of I's and C's. We also did an eye–tracker–study of I's foveal fixations during the directives he gave concerning the tail of the plane.

2.2 Example from Task–Oriented Dialogue: Embedded Problem–Solution

The situation in which the following turn exchange is embedded is as follows: C is just busy building up the wheels–part of the airplane. He cannot figure out which type of bolt

* "Baufix" is a registered trademark of Heinz H. Lorenz GmbH, Geretsried, Germany.

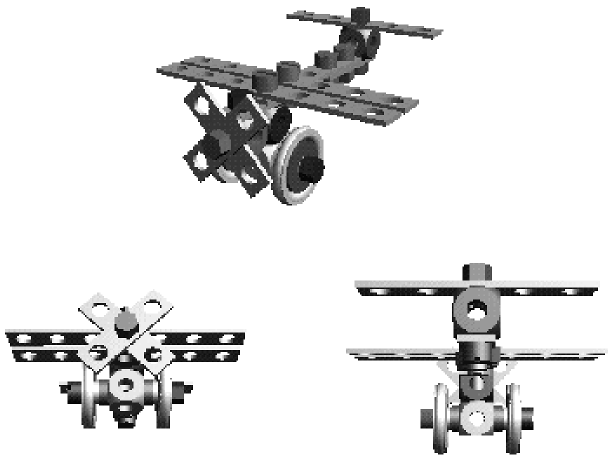


Fig. 1: The "Baufix"–plane seen from various angles

to use on which side. Therefore he issues a test in order to find out about the correct position of the bolts. This can only be achieved if the position of the airplane is correctly fixed, which, therefore has to be accomplished first.

(6)	C:	Also	wenn	die Rückseite	jetzt	zu mir	zeigt,
		Well	if	the top	now	towards me	points,
		das Hinter, *Hinterteil*					
		the bo-, *bottom*					
	I:	*Das Hinterteil*	*zu dir.*				
		The bottom	*towards you.*				
	C:	*des Flugzeugs.*	*Ja, gut*,	dann rechts			
		of the airplane.	*Right, well*,	then on the righthand–side			
		die Eckige		und links			
		the cornered one		and on the lefthand–side			
		die Runde.					
		the round one.					
	I:	*Ja, und*	*links*	*die Runde*			
		Right, and	*lefthand–side*	*the round one.*			
		⟨`pause 3 sec.`⟩ [...].					

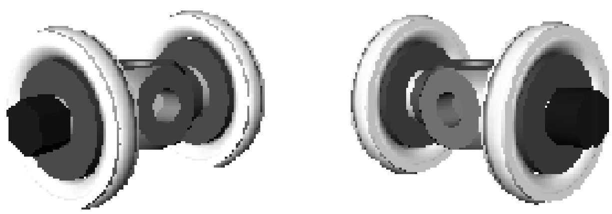


Fig. 2: Positions of right and left bolt respectively in wheels–part of the toy–airplane.

Here the perspective is from the Constructor's point of view, "sitting behind the plane".

The global structure of passage (6) is as shown in Fig. 3: *C* starts to describe the position of the aggregate (already approaching the Platonic state of perfect airplaneness) on his side by producing a discourse particle and an *if*–clause. The *if*–clause is immediately followed by a self–repair. The self–repair introduces a side sequence which has to be completed before *I* can start on locating the bolts. The subject of the side–sequence is a wording–problem. The solution reached via the side–sequence is acknowledged by its initiator, *C*.

	Global structure:	
C:	Adjusting the airplane	*if*–clause
	Embedded structure:	
	Opening of embedded structure:	self–repair
I:	Ratification of adjusting	ellipsis (gapping)
C:	Continuation of adjusting	repair (self or other)
	Closing of adjusting	Acknowledgement
	Closing of embedded structure	
	Continuation of global structure:	
C:	Locating the bolts.	*then*–clause
I:	Acknowledgement	
	⋮	

Fig. 3: Structure of turn exchange in example (6)

The agents have different tasks to accomplish, and remember that *pace* Clark and Wilkes–Gibbs this is a cooperative business. First of all, they have to adjust the airplane on *C*'s side. Then the position of the airplane on *C*'s side has to be grounded, i.e. *I* and *C* must mutually believe that it is such and such and also believe that this is mutually believed.

Given that the position of the bolts on the already adjusted airplane has to be ascertained, this has also to be followed by a grounding phase similar to the one encountered with respect to the adjusting of the airplane. Observe that successfully managing the airplane adjusting phase is a precondition for locating the bolts. Here we clearly see that agents take advantage of the procedurality of speech: Something produced earlier can be negotiated and fixed up such that things to be produced later can depend on it.

2.3 Adding a Wee Bit More Grain to the Example

I now comment on the different stages of the turn exchange in more detail. The first stage is *C*'s adjusting of the airplane, realized by an *if*–clause and a self–repair. *C* self–repairs *die Rückseite/the top* to *das Hinterteil/the bottom.* One possible reason is that *Rückseite* is polysemous between German *Oberseite* and *Hinterteil.* Compare the following data:

die Rückseite des Hauses	the back of the house
die Rückseite des Käfers	the top of the beetle
die Rückseite des Gartens	the far side of the garden.

In English, which does not have this polysemy, the difference is perhaps best portrayed using *tail* versus *top*. Thus *die Rückseite* would cover the top of the plane as well as its tail.

Now the top does not point towards C, only the tail does. Observe that with *das Hinter, Hinterteil/the bo-, bottom* we have a recursive repair–structure. C first starts the production of *Hinterteil/bottom*, but then he interrupts the production, leaving a fragmentary expression, which is however a genuine word of German, the preposition *hinter/back, rear.*

A possible reason for the interruption of word–production might be what I call here "Freudian hesitation" (the term is not to be taken too seriously): *Hinterteil* in German means *bottom* or *behind*. So it could well be that C regards this as not being appropriate at second thought, and that he therefore hesitates. The repair can be thought of as functioning in this way: First, a suitable reparans is produced, its production involving a repair in itself. This yields *das Hinterteil/the bottom*. Secondly, the reparans can be used to go proxy for the original reparandum *die Rückseite/the top*.

After his self–repair, C *could* continue with the *then*–clause, but this is not what happens. Instead, I explicitly acknowledges the position of the airplane as given by C.

Again we can offer a reason for I's move: According to the principles of mutual responsibility and least collaborative effort (of Clark and Wilkes–Gibbs', see above), the agents will make sure that the airplane is correctly adjusted in order to locate the bolts. I ratifies in order to indicate what has to go into the common ground, or, what amounts to the same thing, which information counts as being public. A closer look at I's ratification reveals that he uses ellipsis of the verb: *Das Hinterteil zu dir./The bottom towards you.* stands for *Das Hinterteil* zeigt *zu dir./The bottom* points *towards you.*. I reformulates C's position description and in doing so, he uses C's repair wording, perhaps conforming to the Garrod/Anderson "output–input–coordination principle". This completes the second stage of the embedded structure, ratification of adjusting.

Surely, after I has ratified the position, everything should be clear and C could go on to indicate the position of the bolts. Obviously, C does not think so, since he continues to discuss the positioning of the plane. This is done by using a repair–construction again: *des Flugzeugs/of the airplane.* Clearly, C's *addendum* seems to be redundant. There are various interpretations possible for it:

(a) C is "emptying" his "production stack", since his production was interrupted by I's perhaps over–cooperative verbal behaviour.

(b) C is self–repairing his own wording *das Hinterteil/the bottom*, thus putting still another repair on top of his repair of a repair.

(c) C is other–repairing I's wording *Das Hinterteil zu dir./The bottom towards you.*.

(a) is not too plausible. Its plausibility would be heightened, if we had an overlap of turns *Das Hinterteil zu dir./The bottom towards you.* and *des Flugzeugs/of the airplane.* Then we could argue that C could not stop his production procecess since it had already been triggered and simply had to come to an end due to the motor mechanisms of speech production. However, there is a clear explanatory difference concerning (b) and (c) in terms of discourse analysis. (b) is continuation of own contribution and (c) is continuation of the contribution of the other. Nevertheless, the syntactic pattern used for the continuation is strikingly similar. Even if C's contribution is underspecified, since we do not know where to glue it onto, we can still say something about its function: C indicates by his extension *des Flugzeugs/of the airplane* that content can be grounded. Hence the role of it is similar to verbatim responses or anticipations in telephone–dialogues or online–dialogues.

This comprises step three. But still, we are not quite done, for C, before he starts to verbalise bolt–location produces *Ja, gut, .../Right, well ....* This does not add anything to the content produced so far, so we can take it as an indication of C's that the embedded sequence starting with his self–repair can be closed. That is, according to Clark and Wilkes–Gibbs' "presentation–acceptance cycle", we have only a suggestion of C's to which I has to react. In any case, the inner frame in (6) is shown as a repair–sequence in its own right. It starts with C's self–repair and ends with his closing signal. We have seen that both parties, I and C, are contributing to the positioning of the airplane.

We will here pause for a while and take stock: What have we got as the result of the joint production? First, an interpretable "patched up" token for the *if*–clause reading as follows:

C:	Also	wenn	*das Hinterteil*	*des Flugzeugs*	jetzt	zu mir	
	Well	if	*the bottom*	*of the airplane*	now	towards me	
	zeigt	...					
	points	...					

Via the joint production, the positioning of the plane on C's side has been accomplished by exchange of *die Rückseite/the top* for *das Hinterteil des Flugzeugs/ the bottom of the plane.* As a consequence of this mutual belief/public information has been accomplished in the following way (I follow Heydrich and Rieser (1995) here, see also the contribution of Heydrich on mutuality and common ground in this volume): I and C have reason to believe that there is the appropriate top–and–tail–airplane–situation on the side of C's. The situation indicates to both of them that the other has reason to believe that there is the appropriate top–and–tail etc. The situation indicates to both of them that the airplane–tail appropriately points towards C. Hence, it is public information that the airplane–tail appropriately points towards C. We can also safely assume, that both, I and C, believe that it is public information that the airplane–tail appropriately points towards C. Observe that the next step, locating the bolts, can only be undertaken, after the position of the airplane was made clear. Nothing much need be said about the remainder of (6). The location of the bolts is provided in C's *then*–clause and we have an acknowledgement of I's, which again will provide mutual knowledge of the right sort, to wit, that every appropriate top–and–tail airplane situation will have a cornered bolt on the righthand side and a round bolt on the lefthand side to fix the wheels of the landing gear.

3 Interference of Discourse Organisation and Syntax–of–Utterance. Some Observations

Preference for Self–correction: In the example (6) above, we have, strictly speaking, two possibilities for looking at *das Hinterteil/the bottom*. We can consider it as a repair or we can take it as an (repaired) apposition to *die Rückseite/the top*. Let us refer to these readings as the "repair–reading" and the "apposition reading" respectively. We will be mainly concerned with the "repair–reading" here. Example (6) shows a feature frequently encountered in two–person discourse, namely, that there is a preference for self–correction which we have in the context of the *if*–clause. From the syntactic point of view the interesting problem is why we intuitively know that *das Hinter, Hinterteil/the bo-, bottom* is designed to substitute *die Rückseite/the top* on the "repair–reading".

A first argument backing the intuition is as follows: In the *if*–clause, the structural positions depending on the verb *zeigt/points*, a nominative NP and a prepositionbal phrase *zu mir/towards me*, are filled. Hence the NP *das Hinterteil/the bottom*, which is also a nominative NP, can only be associated with the nominative NP in the closed clausal frame. Consider the following made up expressions:

(7)	Also wenn	die Rückseite	jetzt	zu mir	zeigt,
	Well if	the top	now	towards me	points,
	des Hinter, Hinterteils$_{\text{[genitive]}}$				
	of the bo-, *bottom*				
(8)	Also wenn	die Rückseite	jetzt	zu mir	zeigt,
	Well if	the top	now	towards me	points,
	dem Hinter, *Hinterteil*$_{\text{[dative]}}$				
	of the bo-, *bottom*				

(7) yields an entirely different reading, expressing that somehow the top of the bottom is at stake. (8), which is even more relevant for the argument, associates two dative NPs with each other, identifying the speaker with the bottom. Hence, the syntax of the expression to be repaired controls the syntactic form of the expression functioning as the repairing phrase. This observation gives a first indication as to how the syntax of phrasal repair can be set up: Just copy the structure of the new complement–daughter into the place of the repaired one. A parsing strategy along these lines was developed in Lisken and Rieser (1990).

Overlap of Verbal Material and Parallel Syntax Constructions in Subsequent Turns: If we look at the response of the instructor's, we see, that he accepted the repair. Confirming that he follows the instructor's directive, he uses, apart from the pronoun of course, the same verbal material as used by the instructor. In addition, we have a sort of gapping construction existing between these two turns: The verb *zeigt/points* can be omitted in the second turn. Note also the strictly parallel construction in the result of the repair, the "virtual" *das Hinterteil jetzt zu mir zeigt/the bottom now towards me points*, and the response *das Hinterteil zu dir/the bottom towards you.*

Systematic Repair Positions and Who Takes Advantage of them: After I's production of *das Hinterteil zu dir/the bottom towards you* there is a systematic position open in the discourse for either

(a) a self–repair of I's or

(b) a confirmation or repair by other agent, i.e. C.

I, following the rule that there is a preference for self–correction, could have added a self–repair in this fashion: *das Hinterteil zu dir des Flugzeugs/the bottom towards you of the airplane.* Instead, C specifies whose tail is at stake. This can be explained by assuming that I simply did not see any necessity to add a self–repair. Indeed, there is not much chance for repairing his own wording, since he copied most of it from C. So, he could only have tried to get in an other–repair, repairing C's already two–fold repair, for which there does not seem to be any necessity either, since obviously, I could indeed follow C's repaired directive. The syntax of C's contribution involves two problems to which we now turn.

Underspecification and Extraposition to the Right: The first thing we can observe in C's contribution *des Flugzeugs/of the airplane* is that there is underspecification in the following sense: First, C could have self–repaired his own repair–of–repair and we have two procedural possibilities here, namely, *das Hinter, Hinterteil des Flugzeugs/the bo-, bottom of the airplane* and *die Rückseite des Flugzeugs/the top of the airplane* in turn repaired to *das Hinterteil des Flugzeugs/the bottom of the airplane.* Secondly, C could have other–repaired I's ratification in the following manner: *Das Hinterteil zu dir. des Flugzeugs/The bottom towards you. of the airplane.* Both "landing sites" above are OK from the syntactic point of view, since we have extraposition to the right in German spoken language, where N–Bar–constructions like *Hinterteil/bottom* or *Rückseite/top* can be modified by extraverted material. Indeed, the general observation with respect to spoken German is that it admits modifiers, adjuncts and complements outside the domain of "their" respective projections. Below you find an example for each, illustrating these extraverted uses.

Modifier:

Jetzt	müssen	ähm	am gelben Würfel	noch	⟨`pause.`⟩	mhm
Now	must	uhm	at the yellow cube	in addition	⟨`pause.`⟩	uhu

die Reifen	befestigt werden	⟨`pause.`⟩	des Flugzeugs.
the tires	fastened be	⟨`pause.`⟩	of the plane.

Here the $\mathrm{NP}_{\mathrm{gen}}$ *des Flugzeugs/of the plane* belongs to *Reifen/tires*, i.e. it modifies the N′ *Reifen/tires.*

Adjunct:

Also Du hast doch eben einen Würfel ⟨`par` C: Mhm⟩
Well, you have indeed just now a cube ⟨`par` C: Uhm⟩
angeschraubt mit der roten Schraube.
screwed on with the red screw.

In this example, *mit der roten Schraube/with the red screw* can either be taken as a complement to the finite verb *hast angeschraubt/have screwed on* or as an adjunct to the VP *hast einen Würfel angeschraubt/have a cube screwed on.*

Complement:

Daran ⟨`pause.`⟩ ähm ⟨`pause..`⟩ mußt du mit den Schrauben
At this ⟨`pause.`⟩ uhm ⟨`pause..`⟩ must you with the screws
befestigen zwei Dreierteile ⟨`pause..`⟩.
fasten two threer–parts ⟨`pause..`⟩.

Befestigen/fasten is a transitive verb and needs an obligatory direct object. *Zwei Dreierteile/two threer–parts* is extraverted to the right and posited outside the first sentence boundary.

We now investigate German extraversion to the right in a more systematic fashion.

4 Looking at German Extraversion to the Right

4.1 Empirical Regularities

In trying to find out which regularities simple extraversions in German obey, we take a simpler "within–turn" example in order not to be impeded by the intricacies of joint production. It runs like this:

(9) Also Du hast doch eben einen Würfel
Well, you have indeed just now a cube
⟨`par` C: Mhm⟩ angeschraubt mit der roten Schraube.
⟨`par` C: Uhu⟩ screwed on with the red screw.

The situation is as depicted in Fig. 4.

We can neglect here *C*'s back–channel behaviour ⟨`par` C: *Mhm*⟩ overlapping *I*'s *einen Würfel.* Now the basic intuition with respect to (9) seems to be that we indeed have a sentence–like entity *Also Du hast doch eben einen Würfel angeschraubt/Well, you have just now a cube screwed on* and a prepositional phrase placed outside it, *mit der roten Schraube/with the red screw* which belongs closely to the verb *angeschraubt/screwed on.*

Indeed, we can regard the prepositional phrase as a complement of the verb or an adjunct to the whole verb–phrase since we also can have the more written German–like version

(10) I: Also Du hast doch eben einen Würfel
Well, you have indeed just now a cube
mit der roten Schraube angeschraubt.
with the red screw screwed on.

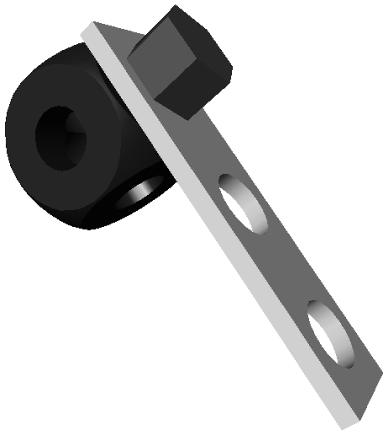


Fig. 4: Screwed on cube referred to in turn (9).

The relation existing between the sentence and the extraverted material does not seem to depend on particles and conjunctions like *also/well*, *doch/nevertheless* or *eben/just now*, so we safely can eliminate them. Similarly, wee need not use the German perfect tense with its finite and infinite part *hast/have ... angeschraubt/ screwed on* in order to get the extraversion effect; the verb in present tense consisting of a separable particle *an/on* and the remainder *schrauben/screw* will serve the same purpose. Finally, the definite NP seems to be either a definite description or an anapher. Both introduce problems of their own and do not seem to be tied up with extraversions. Hence we choose the more tractable (11):

(11)	I:	Du	schraubst	einen Würfel	an	mit einer roten Schraube.
		You	screw	a cube	on	with a red screw.

In order to find out the semantic backing for extraversions, let us have a look at the possible readings of *anschrauben/screw* on by specifying different case–frames of *anschrauben/screw on*. We can have the following with (iii) entailing (ii) and (ii) entailing (i):

(12)	(i)	?anschrauben(*Agens*)	screw–on(*AGENT*)
	(ii)	anschrauben(*Agens*, *Objekt*)	screw–on(*AGENT*, *OBJECT*)
	(iii)	anschrauben(*Agens*, *Objekt*, *Instrument*)	screw–on(*AGENT*, *OBJECT*, *INSTRUMENT*)

(12 i) is perhaps a bit doubtful. Turning now to the syntactic structure of (11), we see that, just as in the more complicated joint production (6), we have an underspecified structure and several possible resolutions of it. Below we give first the underspecified structure (13 i), followed by its intuitively admissible resolutions (13 ii, iii, and iv):

(13) (i) $[_{S}$ $[_{S}$ NP_{nom} $[_{VP}$ V_{fin} NP_{acc} $V_{particle}$]]
Du schraubst einen Würfel an
You screw a cube on
mit–PP]
mit einer roten Schraube.
with a red screw.

(ii) $[_{S}$ NP_{nom} $[_{VP}$ V_{fin} NP_{acc}
Du schraubst einen Würfel
You screw a cube
mit–PP $V_{particle}$]]
mit einer roten Schraube an.
with a red screw on.

(iii) $[_{S}$ NP_{nom} $[_{VP}$ V_{fin} $[_{NP_{acc}}$ Det [N +
Du schraubst einen Würfel
You screw a cube
mit–PP]] $V_{particle}$]]
mit einer roten Schraube an.
with a red screw on.

(iv) $[_{S}$ $[_{S}$ $[_{NP_{nom}}$ Pron + *mit*–PP] $[_{VP}$
Du mit einer roten Schraube
You with a red screw
V_{fin} NP_{acc}]] $V_{particle}$]
schraubst einen Würfel an.
screw a cube on.

(13 i) is straightforward. Here we simply do not know where the *mit*–PP has to go to. Intuitively plausible attachment is provided by the following resolutions: (13 ii) gives you the reading, called "instrument reading", where the cube is fastened with a screw. (13 iii) says that the cube has a screw somewhere, therefore *with a red screw* is modifying cube, hence we dub this reading "modifier reading". (13 iv) is perhaps a relatively odd reading, named "subject reading", saying that it is the person addressed who is the holder of the screw. Topicalized versions and question versions of (13 i) clearly indicate, however, that the three readings, "instrument reading", "modifier reading", and "subject reading" exist:

Topicalized versions:

(ii') Mit einer roten Schraube schraubst du einen Würfel an.
With a red screw screw you a cube on.
("instrument reading")

(iii') Einen Würfel mit einer roten Schraube schraubst du an.
A cube with a red screw screw you on.
("modifier reading")

(iv') Du mit einer roten Schraube bist es, der einen Würfel
You with a red screw is it, who a cube

anschraubt.
screws on.
("subject reading")

Questions:

(ii") Ist es mit einer roten Schraube, daß du einen Würfel
Is it with a red screw that you a cube

anschraubst?
screws on?
("instrument reading")

(iii") Ist es ein Würfel mit einer roten Schraube, den du anschraubst?
Is it a cube with a red screw which you screw on?
("modifier reading")

(iv") Bist du mit einer roten Schraube es, der einen Würfel
Is you with a red screw it who a cube

anschraubt?
screws on?
("subject reading")

Now the crucial question from the point of view of discovering syntactic regularities is this: Which possibilities do we have to extravert syntactic material to the right and which readings do we get in doing so? Let us first stick to the original constituents, *einen Würfel/a cube*, *mit einer roten Schraube/with a red screw*, and *Du/You*.

Extraversions of Maximal Projections Our first case is (13 i) with extraversion of the *mit*–PP, referred to below as (14 i). Next we can exchange the NP_{acc} and the *mit*–PP positions, which yields (14 ii). (14 iii) and (14 iv) demonstrate that we can extravert both complements and exchange their positions in the extraversion slot, although with different consequences for the syntax and generating different entailments.

(14) (i) $[_{S}\ [_{S}\ NP_{nom}\ [_{VP}\ V_{fin}\ NP_{acc}\ V_{particle}\ ]]$
Du schraubst einen Würfel an
You screw a cube on
mit–PP]
mit einer roten Schraube.
with a red screw.

(ii) $[_{S}\ [_{S}\ NP_{nom}\ [_{VP}\ V_{fin}$ *mit*–PP
Du schraubst mit einer roten Schraube
You screw with a red screw
$V_{particle}\ ]]\ NP_{acc}\ ]$
an einen Würfel.
on a cube.

(iii) $[_{S}\ [_{S}\ NP_{nom}\ [_{VP}\ V_{fin}\ V_{particle}\ ]]$
Du schraubst an
You screw on
mit–PP NP_{acc}]
mit einer roten Schraube einen Würfel.
with a red screw a cube.

(iv) $[_{S}\ [_{S}\ NP_{nom}\ [_{VP}\ V_{fin}\ V_{particle}\ ]]\ NP_{acc}$
Du schraubst an einen Würfel
You screw on a cube
mit–PP]
mit einer roten Schraube.
with a red screw.

Now the interesting thing is of course which positions yield which readings: (14 ii) and (14 iii) only have the "instrument reading", whereas (14 iv) has both the "instrument reading" as well as the "modifier reading". We gather from this behaviour of (14 iii) and (14 iv) that the N–Bar *Würfel/cube* modified cannot be transferred to the right and, more generally, that *mit*–PP modifiers cannot stand to the left of the host structure modified.

So far, we haven't yet investigated whether we can push NP_{nom} to the right. So let us have a try at that:

(14) (v) * $[_{S}\ [_{S}\ [_{VP}\ V_{fin}\ V_{particle}\ ]]\ NP_{acc}$
Schraubst an einen Würfel
Screw on a cube
mit–PP NP_{nom}]
mit einer roten Schraube du
with a red screw you.

(vi) * $[_{S}\ [_{S}\ [_{VP}\ V_{fin}\ V_{particle}\ ]]$ *mit*–PP
Schraubst an mit einer roten Schraube
Screw on with a red screw
$NP_{acc}\ NP_{nom}$]
einen Würfel du
a cube you.

In (14 v and vi) the "*" is well placed. These and similar data show that the NP_{nom} in the role of subject cannot go to the right and cannot be placed outside the clausal boundary. Example (14 vii) reveals that a prenominal modifier from a constituent inside the clause *roten/red* cannot be extraverted to the right:

(14) (vii) * Du schraubst einen Würfel mit einer Schraube an,
You screw a cube with a red screw on,
roten.
$\text{red}_{[\text{dative}]}$.

Summing up, we have the following regularities with respect to linear order: modifiers cannot be separated from their modifies by extraversion, subjects cannot be extraverted and neither can prenominal modifiers leaving their hosts behind.

Next we try to find out how post–nominal modification behaves. Here the idea is to keep our original constituents *einen Würfel/a cube*, *mit einer roten Schraube/with a red screw*, and *Du/You* fixed and to add some modification outside the clause.

Post–nominal Modification "Looking to the Left": Consider the following case:

(15) (i) Du schraubst einen Würfel mit einer roten Schraube an
You screw a cube with a red screw on
vom Heck.
from the tail.

In (15 i) the modifying structure *vom Heck/from the tail* is extraverted. As in our simpler examples (13) above, this again results in underspecification and the resolution thereof gives us all three readings, the "instrument reading", the "modifier reading", and the "subject reading". If we extravert our original constituents, we get the modifier *vom Heck/from the tail* attached to the constituent extraverted. The extraverted constituent in turn bars the modifier from attaching to another complement inside the first sentence boundary, i.e. left from the particle *an/on*:

(15) (ii) Du schraubst einen Würfel an mit einer roten Schraube
You screw a cube on with a red screw
vom Heck.
from the tail.

(15) (iii) Du schraubst mit einer roten Schraube an einen Würfel
You screw with a red screw on a cube
vom Heck.
from the tail.

Due to the modifiers choosing the nearest attachment point, we get either the "instrument reading" (15 ii) or the "modifier reading" (15 iii). As to be expected, relative clauses show the same behaviour as PP–modifiers:

(15) (iv) Du schraubst eine Leiste an mit einer roten Schraube,
You screw a bar on with a red screw
die vom Heck ist
which from the tail is.

However, (15 v, vi and vii) demonstrate that attachment can be forced by agreement behaviour of constituents:

(15) (v) Du schraubst einen Würfel mit einer Schraube an
You screw a cube with a screw on,
die rot ist.
which red is.

(vi) Du schraubst eine Leiste mit einer Schraube an,
You screw a bar with a screw on,
die rot ist.
which is red.

(vii) Du schraubst einen Würfel mit einer Schraube an,
You screw a cube with a screw on,
der rot ist.
which$_{\text{[male]}}$ red is.

(15 v and vi) behave as expected: The relative clause attaches to the nearest complement available, even if, as in (15 vi), there would be the possibility to alternatively select the more distant complement. However, in (15 vii) the extraverted relative clause forces attachment to the more distant complement due to matters of agreement, the relative pronoun, being of male gender, can only go with *Würfel/cube*, which is also male. So there remains the mystery, why we do not have underspecification with (15 vi). Perhaps there is a default rule demanding that in case of a constituent which, due to its congruence features, could in principle attach to several constituents to the left, we have to select the nearest one available. But this could well be a tendency of language use to which we are attuned.

Probing Deeper into Recursion: Syntax does, of course, not consist in giving *ad hoc* descriptions of single examples taken from a corpus. That is why we already investigated one recursive step with respect to extraversion to the right, namely, modifying constructions attaching to some N′ (see the examples in (15) above). In addition, we have to look whether extraverted constructions can attach at some "deeper" level of embedding or at some level "in between". We again use relative clauses and prepositional phrases in the extraverted position.

(16) (i) *Du schraubst eine Leiste, die rot ist mit einer Schraube
You screw a bar, which is red with a screw,
die grün ist an die lang ist.
which is green on which is long.

(ii) Du schraubst Leisten mit Löchern mit Schrauben mit Schlitzen
You screw bars with holes with screws with slits
an die rund sind.
on which round are.

(iii) *Du schraubst einen Würfel mit einem Aufsatz mit einer Leiste
You screw a cube with a prolongation with a bar
mit Löchern zusammen, der rund ist.
with holes together, which round is.

(iv) Du bewahrst die Bahncard mit Versicherungsnummer auf
You keep safe the railway card with the insurance number
mit der Scheckkarte mit Geheimnummer, die registriert ist.
with the cheque–card with the secret number which is registered.

(v) *Du bewahrst die Bahncard mit Versicherungsnummer auf
You keep safe the railway card with the insurance number
mit dem Scheckbuch,
with the cheque–book with the secret number
die registriert ist, which$_{\text{[female]}}$ is registered.

(vi) *Du bewahrst das Scheckbuch mit Versicherungsnummer auf
You keep safe the cheque–book with the insurance number
mit der Bahncard mit Geheimnummer,
with the rail–card with the secret number,
das registriert ist, which$_{\text{[neuter]}}$ is registered.

(vii) Du bewahrst das Scheckbuch von Peter mit Versicherungs-
You keep safe the cheque–book of Peter with the insurance
nummer auf mit der Bahncard mit Geheimnummer,
number with the rail–card with the secret number,
die registriert ist von Maria.
which is registered of Maria.

(viii) *Du bewahrst das Scheckbuch mit Versicherungsnummer auf
You keep safe the cheque–book with the insurance number
mit der Bahncard von Maria mit Geheimnummer,
with the rail–card of Maria with the secret number,
die registriert ist von Peter.
which is registered of Peter.

(ix) *Du schraubst einen Würfel mit einer Schraube an,
You screw a cube with a screw on,
der Löcher hat, die rund ist.
which has holes, which round are.

What can we observe with respect to the examples in (16)? (16 i) tells us that if the position inside the clausal structure is occupied as it is here with relative clauses, an extraversion construction of the same type leads to an ungrammatical expression. (16 ii) follows our newly discovered "proximity–to–the–left"–principle, hence *die rund sind/which round are* goes to *Schlitzen/slits*, which, by the way, is rather counter–intuitive from the world–knowledge point of view. In (iii) there is no nearest attachment point available and *der rund ist/which round is* does not hook up with either *Würfel/cube* or *Aufsatz/prolongation* which both are male. (iv) shows again the "proximity–to the–left"–principle, the nearest attachment position *Geheimnummer/secret number* being outside the clause. (v) to (viii) are relevant for the "proximity–to–the–left"–principle as well: In (v), *die registriert ist/which [female] is registered* cannot go with either *Bahncard* or *Versicherungsnummer*, although both are of female gender. This demonstrates that there is no forcing of attachment here, hence we have a difference to the example (15 vii) above, which it would be worthwhile to investigate in greater depth. (vi) covers essentially the same point, "no forcing of attachment". (vii) seems to be all–right: *von Maria/of Maria* can go with *Bahncard/rail–card* or even with *Geheimnummer/secret number*. However, attachment does not go inside the clause as (viii) and (vii') below show:

(vii')*Du bewahrst das Scheckbuch von Peter mit Versicherungs-
You keep safe the cheque–book of Peter with the insurance
nummer mit der Bahncard mit Geheimnummer,
number with the rail–card with the secret number,
die registriert ist auf von Maria.
which is registered of Maria.

In (vii'), *von Maria* cannot attach to *Bahncard*. In (viii) the clause–external position is occupied and we have *von Peter/of Peter* dangling. (ix) indicates, that attachment to the left cannot be carried out crosswise.

Taking Stock: Extraversion to the Right and Recursion What do we gather from the discussion of the examples under (15) and (16)? In the sequel we have listed our findings:

1. Except the NP_{nom} serving as the subject, all complements attached to the finite verb can be extraverted to the right, individually or together in arbitrary order.
2. Modifying constructions cannot stand to the left of their extraverted modified constructions.
3. Modifiers for the N′–level at the first level of the hierarchy yield different readings, depending on the attachment point.
4. There is a "proximity–to–the–left"–principle with a possible exception concerning "attachment–forcing".
5. Extraverted constituents cannot attach in a crosswise fashion.

All these regularities are to be taken as tentative, of course, since we are at the very beginning of research in this area, but they are very good candidates for regularities, nevertheless.

5 Voting for Procedural Explanations in Syntax and Elsewhere

In Section 3 above we discussed under the heading "Underspecification and Extraposition to the Right" that the construction *des Flugzeugs/of the airplane* in our turn–exchange example (6), repeated here as (21), is a case of extraversion to the right and is underspecified. This led to our lengthy discussion about within–turn extraversion phenomena, which, we think, can be captured in an HPSG–approach fairly easily.

(21) C: Also wenn die Rückseite jetzt zu mir zeigt,
Well if the top now towards me points,
das Hinter, Hinterteil
the bo-, bottom
I: Das Hinterteil zu dir.
The bottom towards you.
C: des Flugzeugs. Ja, gut, dann rechts
of the airplane. Right, well, then righthand–side
die Eckige und links die Runde.
the cornered one and lefthand–side the round one.
I: Ja, und links die Runde
Right, and lefthand–side the round one.
⟨`pause 3 sec.`⟩ [. . .].

How can we model some of what happens in the turns above? Obviously, we have to use two syntax–processes, call them I and C, each invested with generating and parsing facilities. C produces the turn (22)

(22) C: Also wenn die Rückseite jetzt zu mir zeigt, das Hinter,
Well if the top now towards me points, the bo-, bottom.
Hinterteil

I parses C's turn and recognises the repair,

(23) das Hinter, Hinterteil
the bo-, bottom,

roughly along the lines sketched in chapter three under the heading "Preference for Self–correction". Now his discourse model tells him that he can demonstrate to C his understanding of the repaired turn by producing an utterance incorporating the repair

(24) I: Das Hinterteil zu dir.
The bottom towards you.

This utterance is in turn parsed by C. C's discourse model now tells C that he can venture an other–repair and that I's offered N–Bar–construction (remember Clark and Wilkes–Gibbs' *presentation–acceptance cycle*) can be a "landing site" for his genitive–phrase. In this sense the presented construction is, according to one reading, not simply accepted but repaired:

(25) des Flugzeugs
of the airplane.

Again resorting to his discourse model, especially to the relevant constraints for opening up and closing side–sequences, he finds that a closure sign for the side–sequence can now be given. Hence, he produces one:

(26) Ja, gut
Right, well.

He thereby marks the interlude as terminated and starts taking up the interrupted *if–then*–clause again:

(27) dann rechts die Eckige
then righthand–side the cornered one
und links die Runde.
and lefthand–side the round one.

I recognises via his discourse model that the side–sequence has been orderly closed by *Yes, well* and that the discourse continues at the "object–level" again.

In order to explain what goes into the common ground of *C* and *I*, we have to resort to what has been produced by either speaker as well as what has been brought about cooperatively. *C* repaired with respect to his own utterance. *I*'s rephrasing presupposes *C*'s utterance and *C*'s repair. Finally, *C*'s self– or other–repair (depending on the resolution of the underspecification considered) is parasitic upon earlier productions. Granted, it will not be easy at all to implement two syntax–processes in the suggested manner but if we want to start investigating cooperative syntax productions in spoken language, we have at least to try to enter this route.

References

Brennan, S. and Clark, H.: 1996, Conceptual pacts and lexical choice in conversation, *Journal of Experimental Psychology, Learning, Memory, and Cognition* 22(6), 1482–1493

Clark, H. and Brennan, S.: 1991, Grounding in communication, in L. Resnick, J. Levine, and S. Teasley (eds.), *Perspectives on Socially Shared Cognition*, pp 127–149, American Psychological Association

Clark, H. and Wilkes-Gibbs, D.: 1986, Referring as a collaborative process, *Cognition* (22), 1–39

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

Garrod, S. and Doherty, G.: 1994, Conversation, co–ordination and convention: an empirical investigation of how groups establish linguistic conventions, *Cognition* (53), 181–215

Gibbon, D., Kindt, W., Kummert, F., and Rieser, H.: 1995, *Syntaxkoordination von Sprechern im Diskurs*, Antrag an die Deutsche Forschungsgemeinschaft auf Fortsetzung des Sonderforschungsbereichs 360, "Situierte Künstliche Kommunikatoren"

Heydrich, W. and Rieser, H.: 1995, Public Information and Mutual Error, Techn. Rep. 95/11, SFB360, University of Bielefeld, Germany

Lewis, D.: 1969, *Convention. A Philosophical Study*, Harvard University Press, Harvard, MA

Lisken, S. and Rieser, H.: 1990, *Ein inkrementeller Parser zur Analyse von simulierten Reparaturen*, KoLiBri Arbeitsbericht 29, Forschergruppe Kohärenz, University of Bielefeld, Germany

Polzin, T. and Rieser, H.: 1992, Parsing with situation semantics, *Nordic Journal of Linguistics* (14), 141–189

Pomplun, M., Rieser, H., Ritter, H., and Velichkovsky, B.: 1997, Augenbewegungen als kognitionswissenschaftlicher Forschungsgegenstand, in R. Kluwe (ed.), *Proceedings der KogWis 96, Hamburg*, to appear

Wachsmuth, I. and Jung, B.: 1996, Dynamic conceptualisation in a mechanical–object assembly environment, *Artificial Intelligence Review* 3–4(10), 345–368

Scorekeeping for Conversation-Construction

Atsushi Shimojima, Yasuhiro Katagiri & Hanae Koiso*

ATR Media Integration & Communications Research Laboratories

Abstract

In this paper, we try to lay the foundation for an informational model of human conversations that formally specifies, for each stage of a conversation, what information is or is not made available to conversants through various forms of "cuings" that occur in the conversation. Squarely facing the fact that multiple lines of cuings often co-occur and interact with each other in the course of an actual conversation, we classify, illustrate, and mathematically characterize their interactions on the basis of Barwise and Seligman's general theory of information flow (1997).

A conversation is what conversants construct. Thus, to explain the construction of a conversation is to explain the conversants' behaviors. We may try to do the latter in various ways. With "conversation analysts" (e.g. Sacks et al. 1974), we may appeal to some social conventions that the participants actually attend to and comply with. Or with "discourse analysts" (e.g. Labov and Fanshel 1977), we may appeal to general rules specifying possible sequences of speech acts. Or we might combine two approaches (Traum 1994) or take still another approach.

Whatever path we may take, such an endeavor must involve or presuppose some explanation of *what information is or is not available to the conversants at a given stage of the conversation*. For example, the application of a particular item of the turn-exchange rules would crucially depend on the information available to conversants about the turn-occupancy state at the point; likewise, depending on what information is assumed to be available to a conversant concerning the prior sequence of speech acts, the sequencing rules on speech acts predict different behaviors of the participant. In most cases, theorists manage to correctly *guess* the available information to a participant (by "putting themselves in his or her position") to make specific predictions about the conversant's behaviors. This practice, however, runs the risk of trivializing whatever theory one may have about conversants' behaviors. Any behavioral theory concerning conversations must be augmented by some independent models of how certain information becomes available to conversants at each stage of a conversation.

Information becomes available to conversants in different ways. Some information is directly accessible through perception; some from memory; still other through sheer imagination or random guessing. However, a great amount of information crucial to conversation

* Also with Nara Advanced Institute of Science and Technology.

constructions becomes available by being *conveyed*, or *cued*, by some other facts holding in conversations. (We will discuss a number of examples later.) Our goal is to obtain a model of what we intuitively grasp as information cuings in conversations and formally specify, for each stage of a conversation, what information is and is not made available to conversants through these cuings. Such a model should capture a large, functionally important part of the mechanisms through which information becomes exploitable by conversants for the construction of a conversation.

This paper consists of four sections. Section 1 will give a clearer picture of the intended model by specifying its intended coverage. We will introduce the notion of "meta-communication," as opposed to that of "base-communication," to highlight the class of phenomena to be covered by our model, though largely ignored by the standard semantic studies. Section 2 will start developing an actual model of conversational cuings. We will motivate the conception of information conveyance to be adopted as the basis for our model, and present a mathematical formulation of the conception due to Barwise and Seligman (1997). We will then show how we apply it to characterize the actual instances of cuings, including "dynamic cuings," found in conversations.

The model presented in section 2 is "basic," in the sense that it only covers a single thread of cuing that occurs in conversations. In an ordinary conversation, however, it is a rule rather than an exception that *multiple threads of cuings co-occur and interact with each other*. They may occur *parallely*, *redundantly*, or *complementarily*; a single fact may cue more than one pieces of information *multiply*; a cuing that usually works may be *blocked* by some intervening fact, and may become a *mis-cuing*; one line of cuing may *conflict* with another and may *override* it, while both may *collapse* together. Facing these phenomena squarely, we will devote section 3 to informally classify the nine forms of cuing-interaction mentioned above and section 4 to show how we can formally model each of these forms by slightly extending the basic tools introduced in section 2.

1 Envisioning the Model

Our model aims to capture all kinds of information conveyances, or cuing relations, so far as they are relevant to the construction of a conversation. In this respect, it should be able to provide a formal, unifying framework for the several traditions of empirical works, including: the works of Kendon (1967), Duncan (1974), Beattie et al. (1982), Koiso et al. (1996), and others on information cuings related to turn-exchanges in conversations; the works of Gumperz (1982), Auer and di Luzio (1993), and others on what they call "contextualization cues"; and Geluykens and Swerts (1994), Swerts et al. (1994), Pierrehumbert and Hirschberg (1990), and Nakajima and Allen (1992) on the cuing functions of prosodic features of speech.

In another respect, our project is a rather ambitious generalization of what formal semanticists have been doing on "linguistic" meaning, and as such, it is a partial realization of what Barwise and Perry (1983) in their book (1983). To facilitate the discussion of this point, we introduce the distinction between "base-communication" and "meta-communication" in the kinds of information conveyances found in conversations.

1.1 Meta-Communication

Borrowing an idea from situation semantics (Barwise and Perry 1983) or more originally from Austin (1950), let us assume that typically in uttering a declarative sentence, a speaker describes a particular situation, called the *described situation.* This lets us define a *topic situation* of a conversation as a situation described by some conversant in some utterance during the conversation. Typically, when one talks about a "communication" in a conversation, one means a conveyance of information about a topic situation of the conversation. We call this level of communication a *base-communication* of the conversation. Take, for example, the following brief conversation, originally cited in Goodwin and Goodwin (1993):[1]

Nancy:	**Jeff** made an asparagus pie
	it was s : : **so** [: **goo** : d.
Tasha:	[I love it. °Yeah I love that.
	Nods *Tasha starts to withdraw gaze*

In this conversation, Nancy describes an event, t, in which she ate an asparagus pie made by Jeff. Tasha describes a slightly different situation, t', concerning the relationship between her and Jeff's asparagus pie in general. In our terms, t and t' are topic situations of this strip of conversation, and the conveyances of information about t and t' made by the sentential utterances in the conversation, namely, the conveyances of the information that Jeff made an asparagus pie in t, that it was so good in t, and that Tasha loves Jeff's asparagus pie in t', are base-communications.

However, not all conveyances of information in a conversation are base-communications. They are not even typical. More typical are conveyances of information *about the conversation itself, as opposed to its topic situations.* We call this kind of information conveyances *meta-communications* in conversations. For example, according to the analysis by Goodwin and Goodwin (1993), the cited conversation involves at least the following conveyances of information at the meta-level:

1. Nancy's use of the intensifier "so" conveys the information that some adjective of assessment will follow it.

2. The enhanced prosody of "so" conveys the information that she is highly involved in assessing Jeff's asparagus pie.

[1] Here, boldface indicates some form of emphasis, which may be signaled by changes in pitch and/or amplitude. The left bracket marks the point at which one speaker's talk overlaps the talk of another, and the degree sign ° indicates that the talk following it is spoken with noticeably lowered volume.

3. The nods accompanying Tasha's first utterance convey the information that the statement that she is making agrees with Nancy's earlier assessment of Jeff's asparagus pie.

4. The early start of Tasha's first utterance and the nods accompanying it convey the information that Tasha is highly involved in praising Jeff's asparagus pie in agreement with Nancy.

5. The choice of text in Tasha's second utterance ("Yeah I love that") conveys the information that she still appreciates what is being talked about.

6. The lowered volume and the shift of gaze from Nancy during Tasha's second utterance convey the information that she is now withdrawing from the activity of praising Jeff's pie.[2]

Notice that in each case, the conveyed information is not about the situation, t, in which Nancy ate Jeff's pie, nor about the situation, t', concerning Tasha's attitude toward Jeff's asparagus pies in general. Rather, the information is about the conversation situation itself: it is about the next lexical item to be uttered (item 1), about the intensity of Nancy's involvement in the current activity (items 2), about the direction to which Tasha's first statement is going (item 3), and about the changing intensity of Tasha's involvement in the current activity (item 4, 5, and 6). The items 1–6 are therefore instances of meta-communication in our taxonomy.

As this example already suggests, the conveyance of information at the meta-level can be triggered by a variety of facts holding in a conversation, and these "cuing" facts convey a variety of information about the conversation situation. To give a feel of the diversity of the phenomena, the following table shows a partial list of possible cuing facts and cued information involved in meta-communication, as they are reported in the literature.

[2] Apparently, the cuings in item 5 and 6 go in the opposite directions. According to Goodwin and Goodwin (1993), Tasha is skillfully using this parallel cuing to change the topic of conversation without a blunt termination of the current activity. We will return to this point in section 3.

Cuing Facts	Cued Information
LINGUISTIC FACTORS	ABOUT CONVERSATION-ORGANIZATION
• Text of speech (e.g. choice of texts; parts of speech; particular syntactic constructions)	• Turn exchanges (completion and continuation of a turn; so-called "transition relevance place")
• Power and power shifts in speech	• Boundaries of the information being expressed at the base-level
• Pitch and pitch contour in speech	• Topic shifts and continuation
• Speed and speed shifts (acceleration and deceleration)	• Types of "core" speech acts being engaged (e.g. informing; checking; requesting)
• Rhythmic pattern (e.g. intra- and inter-turn isochrony)	• Global types of activities being engaged (e.g. telling a joke; teasing)
• Code-switches to different "speech styles", dialects, or foreign languages	ABOUT THE CONVERSANTS
EXTRA-LINGUISTIC FACTORS	• Participation status and role constellation
• Posture (e.g. sitting upright; facing the body toward the addressee)	• Attitudes toward the information being expressed, the activity being engaged, and the topic being dealt with
• Body movement (e.g. nodding; clapping hands, rocking)	• Emotional states (e.g. embarrassment)
• Gaze direction and movements (e.g. gaze aversion; eye-contact)	• Gender, ethnicity, and social status

1.2 Comparison to the Semantic Project

We now use the notion of meta-communication to compare the coverage of our intended model and that of the standard semantic studies. We consider dynamic semantics as a sample of rather recent tradition of semantics.

According to Lewis (1979), a conversation c is a game, with a publicized "scoreboard." The scoreboard is constantly updated as c proceeds, by a participant's utterances and other events in c. The information publicized on the board in turn constrains each participant's subsequent actions, by determining their conformity to the participant's local goal and the global conversation rules. Lewis did not make it explicit, but given the aforementioned distinction between base-communication and meta-communication, we can conceptually distinguish two kinds of information thus publicized: information about the topic t of c and information about c itself. Thus at a given stage of c, there are two scoreboards (or two parts of a scoreboard), s_t and s_c, that exhibit the respective kinds of information. Given an event e in c, then, two different updates by e are conceivable: $s_t \overset{e}{\mapsto} {s_t}'$ and $s_c \overset{e}{\mapsto} {s_c}'$.

Historically, dynamic semanticists (Kamp 1981, Heim 1982, Groenendijk and Stokhof 1991) focused on utterances of some expressions of a natural language in c, and studied how they update scoreboards about the topic of c. Thus, their concerns were in the tertiary relation $s_t \overset{e}{\mapsto} {s_t}'$, where updating events e are confined to utterances of some linguistic units and s_t and ${s_t}'$ are scoreboards about c's main topic. Some authors, including Lewis himself and Stalnaker (1978), emphasized that the update potentials of utterances may depend on conversational parameters such as speaker, addressee, referential salience, and point of reference. Thus, they were interested in a slightly different relation, $\langle s_c, s_t\rangle \overset{e}{\mapsto} s_t{}'$. Even in their cases, however, the focus was on the shift from s_t to $s_t{}'$.[3] This confinement of attention to base-communications is only natural, since the project's main concern was *interpretation* of a linguistic unit, namely, the information carried by an utterance by virtue of its syntactic features, and in most cases, the information carried in that way is concerned with the *topic* of the utterance.

Now, the purpose of our project is to capture all forms of information conveyances functionally significant to the construction. Given that, it is imperative that our model cover the conveyances of information about the conversation situation itself (meta-communication), as well as the conveyances of information about the topic situation (base-communication). In the above terms, we need keep track of the shift, $s_c \overset{e}{\mapsto} s_c{}'$ or $\langle s_c, s_t\rangle \overset{e}{\mapsto} s_c{}'$, of the publicized information about the conversational situation c, as well as that of the publicized information about the topic t of c.

In fact, it is at this point where our project is in stark contrast to the standard semantic endeavor. On the one hand, the "meanings" of the syntactic features of an utterance is typically determined by some *conventionalized* semantic rules. Furthermore, a conveyance of information by virtue of the syntactic features of an utterance is typically *intended* by the speaker of the utterance. As a result, the coverage of the standard semantic study of language use has been typically confined to a very special class of information conveyances occurring in conversations: the class of *intentional* and *conventional* conveyances of information done through the *syntactic features* of utterances.

On the other hand, as the previous example from Goodwin and Goodwin (1993) already shows, an information conveyance at the meta-level is often *unintentional*: the cuing to Nancy's heightened involvement in the current activity by her use of prosodically enhances "so" (item 2) is not necessarily intended by her. Furthermore, an information conveyance at the meta-level is often mediated by signals whose meanings do *not* require the existence of *conventionalized* semantic rules: the cuing to Tasha's withdrawal from the current activity by her gaze aversion (item 6) is certainly not based on some conventional rules that determine the meaning of the gaze aversion. Finally, an information conveyance is often mediated by *non-syntactic features* of speech (item 2 again, where the prosodic features of speech plays the role) and even by non-verbal events (item 6 again, where Tasha's gaze shift plays a role).[4]

Barwise and Perry (1983) demanded that "linguistic meaning should be seen within this general picture of a world teeming with meaning" (p. 16), and that "a semantic theory must account for how language fits in to the general flow of information" (p. 45).

[3] Actually, Lewis mentions the possibility of some dog's starting to run during a conversation, and discusses how that event would affect the referential salience of the dog in question. In our taxonomy, his discussion is concerned with an update of s_c, rather than an update of s_t, and thus makes an exception to the present generalization.

[4] This does not mean that an information conveyance at the meta-level can never be of an intentional, conventional, and linguistic kind. So-called "discourse markers" (Schiffrin 1987) such as "oh," "well," and "y'know" seem to convey information at the meta-level, while being conventional kinds of signs that are often used intentionally.

In dealing with meta-communications as well as base-communications, we are forced to view linguistic meaning within a much wider range of information conveyances occurring in a conversation situation, especially, in relation to non-conventional and non-intentional conveyances of information at the meta-level. In this respect, our project is a generalization of standard semantics to the direction that situation semanticists once envisioned.

2 Basic Model

The discussions in the last section naturally lead us to the question, "What is a conveyance of information, anyway?" Or more specifically to our purpose, what is it for a piece of information to be cued in a conversation? Without a prior determination on this point, no claims on the existence or non-existence of particular lines of cuing would be contentful, and no model of conversational cuings would be empirically testable.

2.1 The Concept of Information Flow

Intuitively speaking, whenever a piece of information is said to be conveyed in a conversation, there is some fact, a "cuing fact," in the conversation, and it somehow tells you that some other fact holds in or outside the conversation. But under what conditions does one fact *tell* you that another fact also holds? One natural answer is, "When there is some kind of regularity between two facts that enforces the second fact to hold when the first fact holds." In fact, this is the idea underlying the theories of information flow developed by Dretske (1981), Barwise and Perry (1983), and Barwise and Seligman (1996). Thus, "the transmission of information requires, not simply a set of de facto correlations, but a network of nomic dependencies between condition at the source and the properties of the signal" (Dretske 1981, pp. 76–77); the "systematic constraints are what allow one situation to contain information about another" (Barwise and Perry 1983, p. 94); "information flow results from regularities in a distributed system" (Barwise and Seligman 1996, p. 8).

In addition to the plausibility of this conception on its own right, there are several theoretical and practical merits in adopting it as the basis of our model, with information cuings in conversations viewed as a special case of information flow. First, this conception gives us a handle of developing a empirically testable model of information cuings in conversations. For, under this conception, to claim that there is a cuing relation between two facts is to claim that there is a regular relationship between them, and the latter is something to be established by some statistical analysis of a conversation corpus or of the experimental results. It is no longer in the discretion of a theorist's introspection whether some fact cues another in conversations.

Secondly and perhaps more importantly, this conception lets us nicely separate the issue of information cuings from the issue of how conversants, with their varying cognitive abilities, exploit the cuings in question. In our view, the first is an issue of the environment in which the cognitive agent is placed, and the second is the issue of the interaction between cognitive agents and their informational environments. It is certainly important, and eventually necessary for our purpose, to investigate the latter issue. However, you can hardly talk about the interaction of an agent and the informational environment without knowing what the environment is like.

To see this point more clearly, suppose we adopted some non-objective view of information conveyances, say, the conception that sees a cuing not as the matter of a regularity over the environment, but as the matter dependent on an agent's attention to it and his or her process of "interpreting" it. Then, the investigation of cuings in conversations would become intertwined with a number of issues of the agent's cognitive abilities and processes. This, it seems to us, is analogous to the mistake of trying to understand the ways a person can use a library without investigating what facility the library provides—how many books are owned, how they are arranged in the stacks, what the check-out policy is, and so on. The non-objective view of information cuing would lead to a conflation of the issue of the informational environment and the issue of the agent's interactions with it.

2.2 Barwise and Seligman on Information Flow

Thus, we adopt the conception of information as the matter of regularities governing the environment. Barwise and Seligman (1996) has recently proposed a theory, called "channel theory," in which this conception of information flow is formulated in a mathematically precise manner. We will now present their model of information flow in some detail, in order to build our model of conversational cuings on its basis.

The following three notions, *classification*, *constraint*, and *infomorphism*, are basic building blocks of their theory:

Definition 1 (Classification): A *classification* $\boldsymbol{A} = \langle \mathrm{tok}(\boldsymbol{A}), \mathrm{typ}(\boldsymbol{A}), \models_{\boldsymbol{A}} \rangle$ consists of

1. a set $\mathrm{tok}(\boldsymbol{A})$ of objects to be classified, called the *tokens of* $\boldsymbol{A}$,
2. a set $\mathrm{typ}(\boldsymbol{A})$ of objects used to classify the tokens, called the *types of* $\boldsymbol{A}$,
3. a binary relation $\models_{\boldsymbol{A}}$ between $\mathrm{tok}(\boldsymbol{A})$ and $\mathrm{typ}(\boldsymbol{A})$.

Definition 2 (Constraint): Let $\boldsymbol{A}$ be a classification. A *sequent* in $\boldsymbol{A}$ is a pair $\langle \Gamma, \Delta \rangle$ of sets of types of $\boldsymbol{A}$. We say that Γ *entails* Δ in $\boldsymbol{A}$, written $\Gamma \vdash_{\boldsymbol{A}} \Delta$, iff every token a of $\boldsymbol{A}$ that is of every type in Γ is of at least one type in Δ. If $\Gamma \vdash_{\boldsymbol{A}} \Delta$ then the pair $\langle \Gamma, \Delta \rangle$ is called a *constraint* supported by the classification $\boldsymbol{A}$.

Definition 3 (Infomorphism): An *infomorphism* $f : \boldsymbol{A} \rightleftarrows \boldsymbol{C}$ from $\boldsymbol{A}$ to $\boldsymbol{C}$ is a contravariant pair of functions $f = \langle f^{\hat{}}, f^{\check{}} \rangle$ satisfying the condition:

$c^{f^{\check{}}} \models_{\boldsymbol{A}} \alpha$ iff $c \models_{\boldsymbol{C}} \alpha^{f^{\hat{}}}$

for each token $c \in \mathrm{tok}(\boldsymbol{C})$ and each type $\alpha \in \mathrm{typ}(\boldsymbol{A})$.

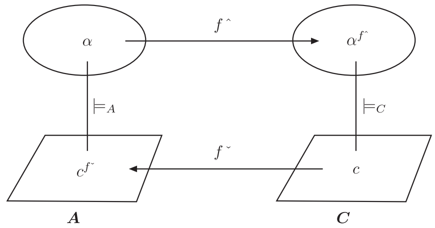


Figure 1

The main function of an infomorphism $f : \boldsymbol{A} \rightleftarrows \boldsymbol{C}$ is to let us express a fact in the classification $\boldsymbol{A}$ as an equivalent fact in the classification $\boldsymbol{C}$. More specifically, if α is a property of the $f^\vee$-value of a token c, then we may take $\alpha^{f^\wedge}$ as the "corresponding" property interpreted as a property of the token c. This is guaranteed by the bi-conditional in the above definition. Thus, intuitively, $c^{f^\vee} \models_{\boldsymbol{A}} \alpha$ can be taken as a fact that the f-value of c is of type α; in contrast, $c \models_{\boldsymbol{C}} \alpha^{f^\wedge}$ is a fact that the token c is of type of having its f-value be of type α. These are equivalent, yet distinct facts. And that we can translate a fact about the value of a token c under some function into an equivalent fact about c itself will have crucial technical importance in Barwise and Seligman's theory.[5]

Definition 4 (Channel): A *channel* $\mathcal{C}$ is an indexed family $\{f_i : \boldsymbol{A}_i \rightleftarrows \boldsymbol{C}\}_{i \in I}$ of infomorphisms with a common codomain $\boldsymbol{C}$, called the *core* of $\mathcal{C}$. The tokens of $\boldsymbol{C}$ are called *connections*; a connection c is said to *connect* the tokens c^{f_i} for $i \in I$. A channel with index set $\{0, \ldots, n-1\}$ is called an *n-ary* channel.

Given that an infomorphism lets us express a fact in its domain classification as a fact in its codomain classification, an indexed family of infomorphisms with a common codomain should let us express a fact in the domain of *each* infomorphism in the family as a fact in the common codomain. To be more precise, let $\boldsymbol{C} = \{f_i : \boldsymbol{A}_i \rightleftarrows \boldsymbol{C}\}_{i \in I}$ be a channel and let c be a particular token in the core classification $\boldsymbol{C}$. Then, for an arbitrary component classification $\boldsymbol{A_i}$, we can express a fact $c^f \models_{\boldsymbol{A_i}} \alpha$ in $\boldsymbol{A_i}$ as the fact $c \models_{\boldsymbol{C}} \alpha^f$ in the core classification $\boldsymbol{C}$.

Combine this idea with the notion of constraint on classifications, or more specifically, with the notion of constraint on the core classification $\boldsymbol{C}$ of the channel $\boldsymbol{C}$. Then we can express the constraints governing the classification relations $\models_{\boldsymbol{A_i}}$ of various component classifications $\boldsymbol{A_i}$ of the channel $\mathcal{C}$ in terms of the constraints on the core classification $\boldsymbol{C}$.

To be more specific, let $\mathcal{C} = \{f_i : \boldsymbol{A}_i \rightleftarrows \boldsymbol{C}\}_{i \in I}$ be an information channel, with $k, l, m, n \in I$. Let a, b, d, g be tokens and $\alpha, \beta, \delta, \gamma$ be types of component classifications

[5] When no confusion is likely, we will suppress the superscripts $^\wedge$ and $^\vee$ for the up- and down-functions in an infomorphism, writing "α^f" and "c^f" for "$\alpha^{f^\wedge}$" and "$c^{f^\vee}$" for example.

$\boldsymbol{A_k}, \boldsymbol{A_l}, \boldsymbol{A_m}, \boldsymbol{A_n}$ of $\mathcal{C}$ respectively. Then, if $\alpha^{f_k} \vdash_{\boldsymbol{C}} \delta^{f_m}$ holds[6], this means that for each token a in A_k, if it is of type α, then each token in A_m connected to a by some connection c in $\boldsymbol{C}$ is of type δ. Also, if $\{f_l\hat{}(\beta), f_m\hat{}(\delta)\} \vdash_{\boldsymbol{C}} f_n\hat{}(\gamma)$, it means that for each pair of tokens b in $\boldsymbol{A_l}$ and d in $\boldsymbol{A_m}$, if b is of type β and d is of type δ, then each token g in $\boldsymbol{A_n}$ connected to b and d by some connection c in $\boldsymbol{C}$ is of type γ. Furthermore, if $f_l\hat{}(\beta) \vdash_{\boldsymbol{C}} \{f_n\hat{}(\gamma), f_k\hat{}(\alpha)\}$, it means that for each token b in A_l, if it is of type β, then for each pair of tokens g in $\boldsymbol{A_n}$ and a in $\boldsymbol{A_n}$, if g and a are connected to b by some connection c in $\boldsymbol{C}$, then either g is of type γ or a is of type α. Thus, a channel can be taken as a mathematical model of a system of constraints governing the distributions of types in various components of a complex system (such as conversations).

Now, if a flow of information is a matter of a constraint, then we should be able to use a channel to model the flows of information that can hold among various components of a complex system. This is the main idea underlying Barwise and Seligman's theory of information flow. Here is their informal characterization of information flow:

> Suppose that the token a is of type α. We say that a's being of type α *carries the information that b is of type β*, relative to the channel $\mathcal{C}$, if a and b are connected in $\mathcal{C}$ and if the translation α' of α entails the translation β' of β in the classification $\boldsymbol{C}$ of the connections of $\mathcal{C}$. (Barwise and Seligman 1996, p. 32.)

For some reason, we do not find a more precise version of this characterization in their book. The passage is specific enough to let us flesh it out in more formal terms, though. We use the auxiliary notion of "proposition" for that purpose:

Definition 5 (Proposition): A *proposition in* a classification $\boldsymbol{A}$ is a triple $\langle a, \alpha, \boldsymbol{A}\rangle$, written $[a \models_{\boldsymbol{A}} \alpha]$, consisting of a token a of $\boldsymbol{A}$, a type α of $\boldsymbol{A}$, and $\boldsymbol{A}$ itself. When $a \models_{\boldsymbol{A}} \alpha$, we sometimes call $[a \models_{\boldsymbol{A}} \alpha]$ a *fact in* $\boldsymbol{A}$.

Then we translate the above passage into the following characterization of information flow:

Definition 6 (Information Flow): Suppose $a \models_{\boldsymbol{A}} \alpha$. The fact $[a \models_{\boldsymbol{A_k}} \alpha]$ is said to *carry the information* $[d \models_{\boldsymbol{A_m}} \delta]$*, relative to* $\mathcal{C}$, iff there is a connection $c \in \text{tok}(\boldsymbol{C})$ such that:

- $c^{f_k} = a$ and $c^{f_m} = d$,
- $\alpha^{f_k} \vdash_{\boldsymbol{C}} \delta^{f_m}$.

In this conception, information-carrying is *veridical*: if a fact $[a \models_{\boldsymbol{A_k}} \alpha]$ carries the information $[d \models_{\boldsymbol{A_m}} \delta]$, then $d \models_{\boldsymbol{A_m}} \delta$. This follows immediately from the fundamental property of infomorphism described in definition 3.

[6] More accurately, this constraint should be written as "$\{\alpha^{f_k}\} \vdash_{\boldsymbol{C}} \{\delta^{f_m}\}$." We are omitting the curly braces for a singleton in describing a constraint.

2.3 Cuings in Conversations

Our fundamental hypothesis is that these notions of channel and information flow are suitable to characterize all cuings, including both base- and meta-communications, that are functionally significant for conversation constructions. The following examples, although not completely worked out, will serve as an adequate indication of how do we go about applying our tools to model the particular instances of cuings in conversations.

Recall Nancy and Tasha's conversation discussed in section 1. Item 2 of Goodwin and Goodwin's analysis claims that the enhanced prosody of Nancy's "so" cues her heightened involvement in the ongoing activity (of assessing Jeff's asparagus pie). To capture the kind of cuings described here, we might posit a channel $\mathcal{C} = \{f_i : \boldsymbol{A}_i \rightleftarrows \boldsymbol{C}\}_{i\in\{0,1\}}$ with the component classification $\boldsymbol{A_0}$ of various units of utterances according to their prosodic features and the component classification $\boldsymbol{A_1}$ of conversants at different times according to their participation status. The connections in $\boldsymbol{C}$ will connect utterances and the utterers at some specific times.

Now let $a \in \mathrm{tok}(\boldsymbol{A_0})$ be the utterance of "so" by Nancy, and $b \in \mathrm{tok}(\boldsymbol{A_1})$ be Nancy at the time when she makes a. Let $\alpha \in \mathrm{typ}(\boldsymbol{A_0})$ be the type of prosody that a has, and $\beta \in \mathrm{typ}(\boldsymbol{A_1})$ be the type of heightened participation that b is in. Then, $a \models_{\boldsymbol{A_0}} \alpha$. In our model, the claim in item 2 is translated to the claim that the fact $[a \models_{\boldsymbol{A_0}} \alpha]$ cues $[b \models_{\boldsymbol{A_1}} \beta]$ relative to $\mathcal{C}$. That is, there is a connection c in $\boldsymbol{C}$ that connects a and b and the constraint $\alpha^{f_0} \vdash_{\boldsymbol{C}} \beta^{f_1}$ holds in $\boldsymbol{C}$. Here, the existence of c that connects a and b simply means the fact that the utterance a is made by Nancy b at a particular time. The constraint $\alpha^{f_0} \vdash_{\boldsymbol{C}} \beta^{f_1}$ is roughly equivalent to saying that whenever an utterance has the enhanced prosodic feature α, the utterer is in the heightened participation status β. Our claim is that these two conditions correctly captures the content of item 2.

Depending on the class of cuings that one wants to model, one need equip one's channel with different sets of component classifications and different kinds of connections for our channel. To capture the cuings triggered by the textual features of utterances, for example, one may want the classification of utterances according to their textual features, along with another classification that represents the sort of things cued by them. To obtain a subtler model of the cuings triggered by prosody of speech, one may want three different classifications of utterances for their power, pitch, and speed, rather than a single classification of utterance prosody ($\boldsymbol{A_0}$ above). Also, one may want the classification of various units of utterances according to the global or local speech acts performed by them, that of hand movements according to their trajectories and speeds, that of conversants at different times according to their belief states, or that of turn-exchange states at different times according to their occupancy status.[7]

With the conception of information flow in definition 6, we can also capture what

[7] For each classification thus posited, one may assign various kinds of objects as its types: real numbers for pitch, power, and speed of utterances, sets of quadruples of real numbers for trajectories, sets of possible worlds for belief states, situation types or infons in situation theory for turn-occupancy status, and so on. The notion of classification is entirely general, and allows any set of objects as the type set for a classification, scientifically sophisticated or not.

may be called "dynamic cuings" in conversations. There occur a great number of events during a conversation. A goat may come into the room (Stalnaker 1978) and a dog may jump up (Lewis 1979) during the conversation. Less dramatic examples are movements by conversants such as inhalation and exhalation, change of gaze directions, iconic and non-iconic gestures, and utterances of grammatical or ungrammatical texts. In many instances of these conversational events, it is possible to tell, either predictively and retrospectively, the event's outcome from its initial condition and the features of the event itself. We call the information conveyance involved in such a case a *dynamic cuing*.

Our framework accommodates dynamic cuings in the following way. First, we assume that the component classifications of our channel are divided into (a) the "state" classifications that classify various states in conversations (such as turn-occupancy states, conversants' emotional states, and their participation status at different times) and (b) the "event" classifications that classify various events occurring in conversations (such as the ones cited above). Secondly, we assume that for each *state* classification $\boldsymbol{A_j}$ for our channel $\mathcal{C} = \{f_i : \boldsymbol{A}_i \rightleftarrows \boldsymbol{C}\}_{i \in I}$, there are a pair of special infomorphisms $in_j : \boldsymbol{A_j} \rightleftarrows \boldsymbol{C}$ and $out_j : \boldsymbol{A_j} \rightleftarrows \boldsymbol{C}$. Then we can characterize a dynamic cuing in the following way:

Definition 7 (Dynamic cuing): Let $\boldsymbol{A_k}$ and $\boldsymbol{A_m}$ be state classifications and $\boldsymbol{A_l}$ be an event classification. Suppose there is a connection c in $\boldsymbol{C}$ such that $c^{in_k\check{}} = a$, $c^{f_l\check{}} = b$, and $c^{out_m\check{}} = d$.

- (Case A) Suppose $a \models_{\boldsymbol{A_k}} \alpha$. The fact $[a \models_{\boldsymbol{A_k}} \alpha]$ cues the information $[d \models_{\boldsymbol{A_m}} \delta]$ *dynamically*, relative to $\mathcal{C}$, iff $\alpha^{in_k\hat{}} \vdash_{\boldsymbol{C}} \delta^{out_m\hat{}}$.
- (Case B) Suppose $b \models_{\boldsymbol{A_l}} \beta$. The fact $[b \models_{\boldsymbol{A_l}} \beta]$ cues the information $[d \models_{\boldsymbol{A_m}} \delta]$ *dynamically*, relative to $\mathcal{C}$, iff $\beta^{f_l\hat{}} \vdash_{\boldsymbol{C}} \delta^{out_m\hat{}}$.

This characterization of dynamic cuing is a direct application of the general idea of dynamic information flow (Barwise and Seligman 1996) to conversational cuings. Intuitively, if there is a connection c such that $c^{in_k\check{}} = a$, $c^{f_l\check{}} = b$, and $c^{out_m\check{}} = d$, this means that a is an initial state for the event b that results in a final state d. Thus, Case A is where the fact $[a \models_{\boldsymbol{A_k}} \alpha]$ about the initial state a of the event b carries the information $[d \models_{\boldsymbol{A_m}} \delta]$ about the outcome d of the event b, and Case B is where the fact $[b \models_{\boldsymbol{A_l}} \beta]$ about the event b itself carries the information $[d \models_{\boldsymbol{A_m}} \delta]$ about the outcome d.[8]

Unfortunately, we do not have space to fully discuss many interesting examples of conversational cuings captured in this definition. To list a few, Duncan (1974), Beattie et al. (1982), and Koiso et al. (1996) study the features of an utterance that indicate whether the current speaking turn ends with the utterance or still continues after it. The relevant features are the pitch, the power, and the choice of a lexical item at the end of the utterance in question, and they dynamically cue an outcome of the utterance (whether the current turn ended or still continues). In a similar vein, Sacks, Schegloff, and Jefferson (1974)

[8] Of course, it is possible that the facts $[a \models_{\boldsymbol{A_k}} \alpha]$ and $[b \models_{\boldsymbol{A_l}} \beta]$ *work together* to cue the information $[d \models_{\boldsymbol{A_m}} \delta]$ and it would a special case of "complementary cuing" we characterize later in section 4.

points out that it is crucial for smooth turn-exchanges that a hearer can project, from various features of an utterance, the next possible point of turn-shift (so-called "transition relevance place") before the utterance actually reaches the point. For a rather different kind of application of definition 7, consider the classifications $\boldsymbol{A_k}$ and $\boldsymbol{A_m}$ above to be a single state classification, say $\boldsymbol{G}$, which classifies the common-ground (or "t-scoreboard" defined in section 1.2) at different times of conversations. Then, we can talk about the regularities from the initial condition of the common-ground to the effect of an utterance event on the common-ground due to the event's particular features. We conjecture that this would let us embed the works in dynamic semantics in our general framework.

3 Interactions of Cuings in Conversations

So far, we have found that with suitable choices of component classifications and of connections, the concept of information carrying introduced above can be applied to model the simple form of cuings occurring in conversations, including dynamic cuings. To obtain a realistic view of the class of information available to conversants through cuings in conversation, however, it is not enough to posit a unary or binary channel that only captures a single route of cuings in conversations. Rather, we have to conglomerate a number of component classifications into a single channel to capture the *interactions of multiple threads of cuings during a conversation.* This task is not as easy as it may first appear, mainly because of a number of rather intricate forms of cuing interaction. In this section, we classify and illustrate the nine intricate forms of cuings interactions found by the present authors or reported in the literature.

A. Parallel Cuing: different facts in a conversation convey different pieces of information *parallely.*

B. Redundant Cuing: different facts in a conversation convey the same piece of information *redundantly.*

C. Multiple Cuing: a single fact in a conversation conveys *multiple* pieces of information.

D. Complementary Cuing: multiple facts cue a piece of information in combination, while they do not do so separately.

Example 1: parallel cuing. Recall items 5 and 6 in the example discussed in section 1 (Goodwin and Goodwin 1993). There, the text of Tasha's statement cues her appreciation of the activity of praising Jeff's pies, while her gaze direction and the volume of her voice cues that she is no longer involved in the activity as before. Tasha seems to skillfully use this parallel cuing to propose the change of topic or activity without abruptly terminating the activity initiated by her co-participant.

Example 2: redundant cuing. According to the analysis in Koiso, Shimojima, and Katagiri (1997), a deceleration of speech that occurs in information-giving utterances in

Japanese conveys the information that a new unit of information starts at that point, while an acceleration cues that there is no opening of an information unit. This means that the opening or non-opening of an information unit is often redundantly cued, since in most cases, it is also cued by the textual features (such as the opening or non-opening of new sentences or clauses) of the utterances in question. In fact, redundant cuings are very common in conversations, working as the "fail-safe" device for conveyance of information (Erickson and Schultz 1982).

Example 3: multiple cuing. According to Couper-Kuhlen (1991), when the speech rate of a particular turn-sequence is significantly greater or slower than those of the surrounding sequence, it means that the sequence in question is a "side sequence," namely, a sequence engaged in an activity (typically the repair of some communication problem) subordinate to the main activity of the conversation. Couper-Kuhlen also claims that if the sequence is accelerated rather than decelerated, it means that the subordinate activity in question is something urgent, such as the repair of a serious communication problem that potentially damages some conversant's "face." Thus, the single fact of an accelerated turn-sequences indicates two pieces of information.

Example 4: Complementary. Recall item 4 in the analysis of Nancy and Tasha's conversation discussed in section 1. There, the early start of Tasha's first utterance and the nods accompanying it seem to *work together* to convey the information that Tasha is highly involved in praising Jeff's asparagus pie in agreement with Nancy (item 4). Neither the early start not the nods, taken by itself, seems to cue the Tasha's heightened involvement strongly enough.

Example 5: Complementary. According to Erickson and Shultz (1982) and Auer (1993), abrupt changes in the power and pitch of speech, in the speaker's posture, and in the frequency of accompanying eye-contacts convey the information that the speaker is engaged in a new type of activity. It seems that the changes in more than one of these parameters collectively cue the change of activity. The change in no single parameter cues it strongly enough.

These forms of cuing involves two or more "concurrent" lines of cuings in different configurations. In particular, the concurrent lines of cuings involved in a parallel, redundant, or multiple cuing are *independent* in its cuing force—the holding of each as a cuing line does not require the presence of the other line of cuing. In contrast, a *complementary* cuing is a case in which two cuing facts are involved without making independent lines of cuing. The question is how we differentiate the complementary cuing from the cases of concurrent cuings, especially from redundant cuings. What is it for two facts to *work together* to convey a piece of information? How should we understand the *contribution* of each fact in the collaboration?

E. Cuing Blockage: a fact in a conversation that *normally* conveys a piece of information does not do so in the presence of some other fact in the conversation.

F. Mis-Cuing: a line of cuing that *normally* conveys accurate information conveys misinformation in certain circumstances.

G. Cuing Conflict: two facts in a conversation convey incompatible pieces of information.

H. Cuing Collapse: two line of cuings occur, and both lines of cuing cease to convey the piece of information that they normally convey.

I. Cuing Override: two line of cuings occur, and only one line of cuing ceases to convey the piece of information that it normally conveys.

Example 9: Blockage. As we mentioned, Koiso, Shimojima, and Katagiri (1997) claim that a deceleration of speech rate in information-giving utterances cues the opening of an information unit. They also reported that this cuing is blocked if the deceleration is exceptionally great in degree, or it is preceded by a filler, or it is preceded or succeeded by a long pause. In such a context, we simply think the speaker is stammering, rather than opening an information unit.

Example 10: Mis-cuing. When a communication problem occurs in a conversation, conversants typically initiates a repair and then actually repair the deficiency (Sacks, Schegloff, and Jefferson 1974). According to Couper-Kuhlen (1991), if the turn for repair initiation or actual repair is rhythmically integrated with the previous turn, it normally means that the communication problem being addressed is a simple, acoustical problem (such as the occurrence of a disturbing noise), as opposed to a serious, potentially face-threatening problem (such as the misuse of a technical expression). However, Couper-Kuhlen also shows that the rhythmic integration can mis-cue that the relevant problem is not serious one, while in fact the problem is serious one. Thus, the default cuing by the rhythmic integration can be *abused* to camouflage the seriousness of the problem.

Example 11: Conflict and Override. According to Koiso et al. (1996), a flat pitch and power at the final part of an utterance cues the continuation of the current turn after the utterance in question. The data show that this cuing sometimes conflicts with, and is overridden by the use of a verb in the imperative mood in the same place, which cues the end of the current turn.

Example 12: Conflict and Collapse. In contrast, the same data (Koiso et al. 1996) show that the cuing to a turn-continuation by the use of an adverb and the cuing to a turn-end by a decrease of the power of speech collapse and the message becomes equivocal, when both occur at the same place of an utterance.

All of these forms of cuings are instances of what may be called "default cuings," where a line of cuing that *normally* conveys accurate information may or may not work in some exceptional circumstances. The main challenge is to specify the sense in which a line of cuing that *normally* conveys accurate piece of information, while allowing the possibility that it may be blocked in some *exceptional* circumstances.

4 Modeling the Cuing Interactions

We will see in this section that the basic notions introduced in section 2 are sufficient to model the first four forms of cuing interactions (A–D) described in the last section, while modeling the last five forms of cuing interactions (E–I) requires an extension of our tool kit with the notion of "refinement" (Barwise and Seligman 1997).

4.1 Cuing Interactions: A–D

Let $\mathcal{C} = \{f_i : \boldsymbol{A}_i \rightleftarrows \boldsymbol{C}\}_{i \in I}$ be an information channel, with $k, l, m, n \in I$. Let a, b, d, g be tokens and $\alpha, \beta, \delta, \gamma$ be types of component classifications $\boldsymbol{A_k}, \boldsymbol{A_l}, \boldsymbol{A_m}, \boldsymbol{A_n}$ of $\mathcal{C}$ respectively. (We will assume this setting for all the definitions that follow.)

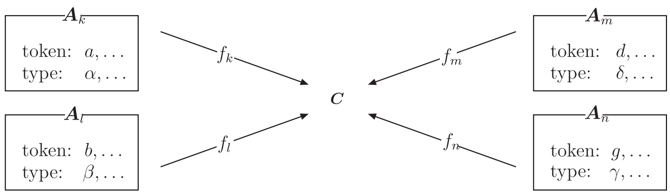


Figure 2

Definition 8 (Parallel, redundant, and multiple cuing): Suppose the fact $[a \models_{\boldsymbol{A_k}} \alpha]$ carries the information $[d \models_{\boldsymbol{A_m}} \delta]$ and the fact $[b \models_{\boldsymbol{A_l}} \beta]$ carries the information $[g \models_{\boldsymbol{A_n}} \gamma]$, relative to $\mathcal{C}$.

1. The facts $[a \models_{\boldsymbol{A_k}} \alpha]$ and $[b \models_{\boldsymbol{A_l}} \beta]$ cue $[d \models_{\boldsymbol{A_m}} \delta]$ and $[g \models_{\boldsymbol{A_n}} \gamma]$ *parallely*, relative to $\mathcal{C}$, iff $[a \models_{\boldsymbol{A_k}} \alpha] \neq [b \models_{\boldsymbol{A_l}} \beta]$ and $[d \models_{\boldsymbol{A_m}} \delta] \neq [g \models_{\boldsymbol{A_n}} \gamma]$.

2. The facts $[a \models_{\boldsymbol{A_k}} \alpha]$ and $[b \models_{\boldsymbol{A_l}} \beta]$ cue $[d \models_{\boldsymbol{A_m}} \delta]$ *redundantly*, relative to $\mathcal{C}$, iff $[a \models_{\boldsymbol{A_k}} \alpha] \neq [b \models_{\boldsymbol{A_l}} \beta]$ and $[d \models_{\boldsymbol{A_m}} \delta] = [g \models_{\boldsymbol{A_n}} \gamma]$.

3. The fact $[a \models_{\boldsymbol{A_k}} \alpha]$ cues $[d \models_{\boldsymbol{A_m}} \delta]$ and $[g \models_{\boldsymbol{A_n}} \gamma]$ *multiply*, relative to $\mathcal{C}$, iff $[a \models_{\boldsymbol{A_k}} \alpha] = [b \models_{\boldsymbol{A_l}} \beta]$ and $[d \models_{\boldsymbol{A_m}} \delta] \neq [g \models_{\boldsymbol{A_n}} \gamma]$.

Note that the characterizations of parallel and redundant cuings above do not require the tokens a and b of the cuing facts $[a \models_{\boldsymbol{A_k}} \alpha]$ and $[b \models_{\boldsymbol{A_l}} \beta]$ to be temporarily concurrent. In our framework, temporarily divergent facts can still make parallel and redundant cuings (as we desired in section 3).

Contrast these case of "concurrent" cuings with the following case of "combinatorial cuing":

Definition 9 (Combinatorial cuing): Suppose $a \models_{\boldsymbol{A_k}} \alpha$ and $b \models_{\boldsymbol{A_l}} \beta$. The facts $[a \models_{\boldsymbol{A_k}} \alpha]$ and $[b \models_{\boldsymbol{A_l}} \beta]$ cue the information $[d \models_{\boldsymbol{A_m}} \delta]$ *in combination*, relative to $\mathcal{C}$, iff:

- there is a connection c in $\boldsymbol{C}$ that connects a, b, and d,

- $\{\alpha^{f_k}, \beta^{f_l}\} \vdash_{\boldsymbol{C}} \delta^{f_m}$.

The combinatorial cuing and the redundant cuing are conceptually different in that the latter implies that each of the involved facts $a \models_{\boldsymbol{A_k}} \alpha$ and $b \models_{\boldsymbol{A_l}} \beta$ cues the information $d \models_{\boldsymbol{A_m}} \delta$ while the former has no such implication.[9]

In fact, we can characterize what we called "complementary cuing" as a case of combinatorial cuing in which neither fact involved in the cuing cues by itself.

Definition 10 (Complementary cuing): Suppose $a \models_{\boldsymbol{A_k}} \alpha$ and $b \models_{\boldsymbol{A_l}} \beta$. The facts $[a \models_{\boldsymbol{A_k}} \alpha]$ and $[b \models_{\boldsymbol{A_l}} \beta]$ complement each other to cue the information $[d \models_{\boldsymbol{A_m}} \delta]$, relative to $\mathcal{C}$, iff:

- $[a \models_{\boldsymbol{A_k}} \alpha]$ and $[b \models_{\boldsymbol{A_l}} \beta]$ cue $[d \models_{\boldsymbol{A_m}} \delta]$ in combination, relative to $\mathcal{C}$,
- $\alpha^{f_k} \nvdash_{\boldsymbol{C}} \delta^{f_m}$ and $\beta^{f_l} \nvdash_{\boldsymbol{C}} \delta^{f_m}$.

4.2 Cuing Interactions: E–I

Barwise and Seligman's theory allows more than one channel to be associated with an environment, making the class of information flows holding in an environment relative to the specific channel in focus. The following notion of refinement, due to Barwise and Seligman (1996), is intended to capture the relationship between two channels such that one embodies a stricter system of constraints than the other while being "continuous" with the other in all the other respects.

Definition 11 (Refinement): Let $\mathcal{C} = \{f_i : \boldsymbol{A}_i \rightleftarrows \boldsymbol{C}\}_{i \in I}$ and $\mathcal{C}' = \{g_i : \boldsymbol{A}_i \rightleftarrows \boldsymbol{C}'\}_{i \in I}$ be channels with the same component classifications $\boldsymbol{A}_i$. A *refinement* infomorphism r from $\mathcal{C}'$ to $\mathcal{C}$ is an infomorphism $r : \mathcal{C}' \rightleftarrows \mathcal{C}$ such that for each i, $f_i = r \circ g_i$, that is, the following diagram commutes:

$$\begin{tikzcd} & \boldsymbol{C}' \arrow[dr, "r"] & \\ \boldsymbol{A}_i \arrow[ur, "g_i"] \arrow[rr, "f_i"] & & \boldsymbol{C} \end{tikzcd}$$

Figure 3

The channel $\mathcal{C}'$ is a *refinement of* the channel $\mathcal{C}$ if there is a refinement r from $\mathcal{C}'$ to $\mathcal{C}$.

[9] The reader may well wonder if there is any kind of cuing that stands to the corresponding relationship to the multiple cuing as the combinatorial cuing stands to the redundant cuing. Theoretically, we can define the notion of *distributed cuing* as the case in which:

- there is a connection in $\boldsymbol{C}$ that connects a, d, and g,
- $\alpha^{f_k} \vdash_{\boldsymbol{C}} \delta^{f_m}$ and $\alpha^{f_k} \vdash_{\boldsymbol{C}} \gamma^{f_n}$.

However, we have not yet explored what real phenomena, if any, of conversation cuings can be characterized by means of this notion.

The commutativity of the above diagram dictates that the refined channel $\mathcal{C}'$ and the "de-refined" channel $\mathcal{C}$ behave in exactly the same way so far as the connections and the types that are linked by the infomorphism r are concerned. Yet $\mathcal{C}'$ and $\mathcal{C}$ may behave differently in other connections and types. In particular, a connection in $\boldsymbol{C}'$ that is not the r-value of any connection in $\boldsymbol{C}$ may behave strange, and may become an exception to a constraint that is respected by all connections in $\boldsymbol{C}$. Thus, the following does *not* generally hold:

> For every sequent $\langle \Gamma, \Delta \rangle$ in the core classification of $\mathcal{C}'$, if $r\hat{}(\Gamma) \vdash_{\boldsymbol{C}} r\hat{}(\Delta)$, then $\Gamma \vdash_{\boldsymbol{C}'} \Delta$.

It is in this sense that a refined channel $\mathcal{C}'$ embodies a stricter system of constraint than the "de-refined" channel $\mathcal{C}$ while being continuous with it.

Using this idea, we can characterize default cuings in the following way:

Definition 12 (Default cuing): Suppose $a \models_{\boldsymbol{A_k}} \alpha$ and $b \models_{\boldsymbol{A_l}} \beta$ and that there is a connection $c \in tok(\boldsymbol{C})$ that connects a, b, and d. The fact $[a \models_{\boldsymbol{A_k}} \alpha]$ cues the information $[d \models_{\boldsymbol{A_m}} \delta]$ *in default* of the fact $[b \models_{\boldsymbol{A_l}} \beta]$, relative to $\mathcal{C}$, iff:

- there is a channel $\mathcal{C}^* = \{h_i : \boldsymbol{A}_i \rightleftarrows \boldsymbol{C}\}_{i \in I}$ such that:
  - $\mathcal{C}$ is a refinement of $\mathcal{C}^*$,
  - $\alpha^{h_k} \vdash_{\boldsymbol{C}^*} \delta^{h_m}$,
  - $\text{token}(\alpha^{h_k}) \neq \emptyset$,
- For all $c' \in tok(\boldsymbol{C})$, if $c' \models_{\boldsymbol{C}} \alpha^{f_k}$ and $c' \not\models_{\boldsymbol{C}} \delta^{f_m}$ then $c' \models_{\boldsymbol{C}} \beta^{f_l}$.

The existence of a de-refined channel $\mathcal{C}^*$ in the first main clause of the definition guarantees that by ignoring some proper set of the cases in which a fact of the type α holds, we could consider the constraint from α to δ to hold. Given the assumption that a is connected to d, this means that there is a definite sense in which the fact $[a \models_{\boldsymbol{A_k}} \alpha]$ normally carry the information $[d \models_{\boldsymbol{A_m}} \delta]$ *normally* holds. The second clause says that if the regularity $\alpha^{h_k} \vdash_{\boldsymbol{C}} \delta^{h_m}$ ever fails, it is when a fact of the type β co-occurs. We propose that this correctly capture the case we would describe as "$[a \models_{\boldsymbol{A_k}} \alpha]$ cues $[d \models_{\boldsymbol{A_m}} \delta]$ in default of $[b \models_{\boldsymbol{A_l}} \beta]$."

Note that this definition allows two possibilities: (1) $\alpha^{f_k} \vdash_{\boldsymbol{C}} \delta^{f_m}$ does hold, and the fact $[a \models_{\boldsymbol{A_k}} \alpha]$ genuinely carries the information $[d \models_{\boldsymbol{A_m}} \delta]$ relative to $\mathcal{C}$, and (2) $\alpha^{f_k} \vdash_{\boldsymbol{C}} \delta^{f_m}$ does not hold:

Definition 13 (Cuing survival): Suppose $[a \models_{\boldsymbol{A_k}} \alpha]$ cues $[d \models_{\boldsymbol{A_m}} \delta]$ in default of $[b \models_{\boldsymbol{A_l}} \beta]$. The default cuing by $[a \models_{\boldsymbol{A_k}} \alpha]$ to $[d \models_{\boldsymbol{A_m}} \delta]$ *survives* $[b \models_{\boldsymbol{A_l}} \beta]$ in $\mathcal{C}$ iff $\alpha^{f_k} \vdash_{\boldsymbol{C}} \delta^{f_m}$.

It follows that the default cuing by $[a \models_{\boldsymbol{A_k}} \alpha]$ to $[d \models_{\boldsymbol{A_m}} \delta]$ is a genuine case of information flow if and only if it survives the fact $[b \models_{\boldsymbol{A_l}} \beta]$. Of course, the more interesting case is where a default cuing does not survive, and fails to be a genuine information flow. The cases of mis-cuing and cuing override, collapse, and conflict discussed in section 3 all involve some "blocked" default cuings.

Definition 14 (Overriding and collapse):

1. The cuing by $[a \models_{\boldsymbol{A_k}} \alpha]$ to $[d \models_{\boldsymbol{A_m}} \delta]$ *overrides* the cuing by $[b \models_{\boldsymbol{A_l}} \beta]$ to $[g \models_{\boldsymbol{A_n}} \gamma]$ in $\mathcal{C}$ iff the cuing to $[d \models_{\boldsymbol{A_m}} \delta]$ by $[a \models_{\boldsymbol{A_k}} \alpha]$ survives $[b \models_{\boldsymbol{A_l}} \beta]$ in $\mathcal{C}$, and the cuing to $[g \models_{\boldsymbol{A_n}} \gamma]$ by $[b \models_{\boldsymbol{A_l}} \beta]$ does not survive $[a \models_{\boldsymbol{A_k}} \alpha]$ in $\mathcal{C}$.

2. The cuings *collapse* in $\mathcal{C}$ iff the cuing to $[d \models_{\boldsymbol{A_m}} \delta]$ by $[a \models_{\boldsymbol{A_k}} \alpha]$ does not survive $[b \models_{\boldsymbol{A_l}} \beta]$ in $\mathcal{C}$ either.

Although a blocked default cuing cannot be a genuine information flow, it does not follow that the information conveyed in it is inaccurate. Thus, the case of mis-cuing discussed in section 3 is only a special case of cuing blockage:

Definition 15 (Mis-cuing): Suppose $a \models_{\boldsymbol{A_k}} \alpha$. The fact $[a \models_{\boldsymbol{A_k}} \alpha]$ *mis-cues* $[d \models_{\boldsymbol{A_m}} \delta]$, relative to $\mathcal{C}$, iff:

- the default cuing by $[a \models_{\boldsymbol{A_k}} \alpha]$ to $[d \models_{\boldsymbol{A_m}} \delta]$ is blocked by some fact in some component classification of $\mathcal{C}$,

- $d \not\models_{\boldsymbol{A_m}} \delta$.

We intend to characterize a cuing conflict as a case in which inconsistent propositions are conveyed by two default cuings. But what is it for a set of propositions to be inconsistent within a channel?

Definition 16 (Inconsistency): Let $\Re$ be a set of propositions in component classifications of the channel $\mathcal{C}$. $\Re$ is *inconsistent in* $\mathcal{C}$ iff:

- there is no connection $c \in \mathrm{tok}(\boldsymbol{C})$ such that, for every $[t \models_{\boldsymbol{A_j}} \theta]$ in $\Re$, c connects t and $c \models_{\boldsymbol{C}} \theta^{f_j}$.

- there is a connection $c \in \mathrm{tok}(\boldsymbol{C})$ such that, for every $[t \models_{\boldsymbol{A_j}} \theta]$ in $\Re$, c connects t.

The first clause says that the propositions in $\Re$ never co-occur, which is reasonable as a condition for inconsistency. The second clause excludes the case in which the propositions in $\Re$ never co-occur simply because their tokens are not connected. Without that clause, for example, *any* pair of propositions whose tokens are not connected would become inconsistent. But such a pair of propositions are unrelated, and therefore do not exclude each other. In our conception, such a pair is consistent, rather than inconsistent.

The definition also prevents any set $\Re$ of *facts* from being inconsistent. For, if there is no connection $c \in \mathrm{tok}(\boldsymbol{C})$ that connects t for every $[t \models_{\boldsymbol{A}_j} \theta]$ in $\Re$, then $\Re$ is not inconsistent by definition 16; if there is such a connection c, then from the fundamental property of infomorphism, $c \models_{\boldsymbol{C}} \theta^{f_j}$ for every $[t \models_{\boldsymbol{A}_j} \theta]$ in $\Re$, and hence $\Re$ is not inconsistent.

This notion of inconsistency lets us define cuing conflict in the following way:

Definition 17 (Cuing conflict): The cuing to $[d \models_{\boldsymbol{A}_m} \delta]$ by $[a \models_{\boldsymbol{A}_k} \alpha]$ and the cuing to $[g \models_{\boldsymbol{A}_n} \gamma]$ by $[b \models_{\boldsymbol{A}_l} \beta]$ *conflict*, relative to $\mathcal{C}$, iff:

- $[a \models_{\boldsymbol{A}_k} \alpha]$ cues $[d \models_{\boldsymbol{A}_m} \delta]$ in default of $[b \models_{\boldsymbol{A}_l} \beta]$ relative to $\mathcal{C}$,
- $[b \models_{\boldsymbol{A}_l} \beta]$ cues $[g \models_{\boldsymbol{A}_n} \gamma]$ in default of $[a \models_{\boldsymbol{A}_k} \alpha]$ relative to $\mathcal{C}$,
- The set $\{[d \models_{\boldsymbol{A}_m} \delta], [g \models_{\boldsymbol{A}_n} \gamma]\}$ is inconsistent.

Neither the cuing override nor the cuing collapse implies that the two default cuings involved are conflicting. On the other hand, if two default cuings conflict, then either they collapse or one overrides the other. For otherwise, both cuings would be genuine information flows, and both of the cued propositions would be facts. Since no set of facts are inconsistent, this contradicts the assumption.

In this section, we have only given pairwise characterizations of cuing override, cuing collapse, and cuing conflict. Generalizations into set-wise definitions should be obvious.

5 Conclusion

We have been trying to lay the foundation for an informational model of human conversations that predict, for each stage of a conversation, what information has or has not been available to conversants through various forms of information conveyances in the conversation. We paid special attention to the cases in which multiple lines of cuings interact with each other, and tried to characterize them on the basis of Barwise and Seligman's theory of channels (1996).

We argued that a satisfactory model should generalize the standard semantics study of language use to cover meta-communications, as well as base-communications, that occur in conversations. We pointed out that this generalization requires us to see linguistic meaning as an instance of the much wider variety of information conveyances in conversations, including non-intentional, non-conventional, and non-linguistic kinds.

On the technical side, we found that the notion of channel in Barwise and Seligman's theory (1996) lets us build a model that is (a) general enough to cover various kinds of conversational cuings, including dynamic ones, under the common conception of information flow and (b) fine-grained enough to differentiate the cases of parallel, redundant, multiple, and complementary cuings found in conversations. Furthermore, extending the basic model with the notion of refinement, we could model rather intricate interactions of cuings that involve default cuings, namely, the cases of mis-cuing, cuing blockage, override, collapse, and conflict.

References

Auer, J. C. P.: 1993, Introduction: John Gumperz' approach to contextualization, in J. C. P. Auer and A. di Luzio (eds.), *Contextualization of Language*, John Benjamins Publishing Co., Amsterdam.

Auer, J. C. P. and di Luzio, A.: 1993, *Contextualization of Language*, John Benjamins Publishing Co., Amsterdam.

Austin, J. L.: 1950, Truth, in J. O. Urmason and G. J. Warnock (eds.), *Philosophical Papers*, pp 117–133, Oxford University Press, 1970, Oxford.

Barwise, J. and Perry, J.: 1983, *Situations and Attitudes*, A Bradford Book, The MIT Press, Cambridge, MA.

Barwise, J. and Seligman, J.: 1996, *Information Flow in Distributed Systems*, Unpublished manuscript. A newer version is to be published as Barwise and Seligman (1997).

Barwise, J. and Seligman, J.: 1997, *Information Flow: the Logic of Distributed Systems*, Cambridge Tracts in Theoretical Computer Science, 42, Cambridge University Press, Cambridge, UK.

Beattie, G. W., Cutler, A., and Pearson, M.: 1982, Why is Mrs Thatcher interrupted so often?, *Nature* 300(23/30), 744–747.

Couper-Kuhlen, E.: 1991, *English Speech Rhythm: Form and Function in Everyday Verbal Interaction*, Ph.D. thesis, The University of Zurich.

Dretske, F. I.: 1981, *Knowledge and the Flow of Information*, A Bradford book, The MIT Press, Cambridge, MA.

Duncan, Jr., S.: 1974, On the structure of speaker-auditor interaction during speaking turns, *Language in Society* 2, 161–180.

Erickson, F. and Schultz, J.: 1982, When is a context? Some issues and methods in the analysis of social competence, in J. L. Green and C. Wallat (eds.), *Ethnography and Language in Educational Settings*, pp 147–160, Ablex Press, Norwood, NJ.

Geluykens, R. and Swerts, M.: 1994, Prosodic cues to discourse boundaries in experimental dialogues, *Speech Communication* 15, 69–77.

Goodwin, C. and Goodwin, M. H.: 1993, Context, activity and participation, in J. C. P. Auer and A. di Luzio (eds.), *Contextualization of Language*, John Benjamins Publishing Co., Amsterdam.

Groenendijk, J. and Stokhof, M.: 1991, Dynamic predicate logic, *Linguistics and Philosophy* 14, 39–100.

Gumperz, J. J.: 1982, *Discourse Strategies*, Studies in Interactional Sociolinguistics, Cambridge University Press, Cambridge, UK.

Heim, I.: 1982, *The Semantics of Definite and Indefinite Noun Phrases*, A Garland Series: Outstanding Dissertations in Linguistics, Garland Publishing, 1988, New York.

Kamp, H.: 1981, A theory of truth and semantic representation, in J. A. G. Groenendijk, T. M. V. Janssen, and M. B. J. Stokhof (eds.), *Formal Methods in the Study of Language: Part 1*, pp 277–321, Mathematisch Centrum, Amsterdam.

Kendon, A.: 1967, Some functions of gaze-direction in social interaction, *Acta Psychologica* 26, 22–63.

Koiso, H., Horiuchi, Y., Tutiya, S., and Ichikawa, A.: 1996, The prediction of the termination/continuation of utterance based on some linguistic and prosodic elements, in *The Proceedings of the Tenth Annual Conference of Japanese Society for Artificial Intelligence*, Japanese Society for Artificial Intelligence.

Koiso, H., Shimojima, A., and Katagiri, Y.: 1997, Informational potentials of dynamic speech rate, in *The Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*, Cognitive Science Society, Stanford, CA.

Labov, W. and Fanshel, D.: 1977, *Therapeutic Discourse: Psychotherapy as Conversation*, Academic Press, New York.

Lewis, D. K.: 1979, Scorekeeping in a language game, in R. Bäuerle, U. Egli, and A. von Stechow (eds.), *Semantics from Different Points of View*, pp 172–187, Springer Verlag, Berlin.

Nakajima, S. and Allen, J.: 1992, Prospdy as a cue for discourse structure, in *Proceedings of the International Conference on Spoken Language Processing*, pp 425–428, Banff, Canada.

Pierrehumbert, J. and Hirschberg, J.: 1990, The meaning of intonational contours in the interpretation of discourse, in P. R. Cohen, J. Morgan, and M. E. Pollack (eds.), *Intentions in Communication*, pp 271–311, The MIT Press, Cambridge, MA.

Sacks, H., Schegloff, E. A., and Jefferson, G.: 1974, A simplest systematics for the organization of turn-taking for conversation, *Language* 50(4), 696–735.

Schiffrin, D.: 1987, *Discourse Markers*, Studies in Interactional Sociolinguistics, Cambridge University Press, Cambridge, UK.

Stalnaker, R. C.: 1978, Assertion, in P. Cole (ed.), *Syntax and Semantics, Volume 9: Pragmatics*, pp 315–332, Academic Press, INC., Orlando, FL.

Swerts, M., Bouwhuis, D. G., and Collier, R.: 1994, Melodic cues to the perceived "finality" of utterances, *The Journal of the Acoustical Society of America* 96(4), 2064–2075.

Traum, D. R.: 1994, *A Computational Theory of Grounding in Natural Language Conversation*, Ph.D. thesis, The University of Rochester, Rochester, NY.

The Common Ground as a Dialogue Parameter

Henk Zeevat

1 Introduction

This paper tries to define a central notion in the semantics of dialogues: the common ground between the speaker and hearer and its evolvement as the dialogue proceeds. The starting point is the theory of pragmatics introduced by Stalnaker in Stalnaker 1978. Here implicatures arise as the preconditions of certain speech acts and presuppositions are defined as the shared assumptions of speaker and hearer. This theory makes the common ground the central notion in understanding speech acts and presuppositions and makes it the parameter which controls decisions of the speaker about his communicative course of action and of the hearer in deciding what to make of the speaker's contribution. In the theory, the common ground is therefore one of the starting points for the explanation of linguistic behaviour and for understanding interaction. In this paper, I try to apply this idea in a characterisation of speech acts by stating (epistemic) preconditions on the common ground for their use, by stating their guaranteed contribution and by indicating the moves for the other party that are available after it.

The paper is innovative in making common grounds a special kind of information states and in making these the basis of an update system. They will not only have facts, but will also have opinions about the beliefs of the speaker and the hearer. I take this to be the crucial step: without it there is not enough expressive power to define which information states are common grounds, the characterisation of the speech acts is approximative only and it is not possible to model conflict. And absence of conflict makes it impossible to apply update semantics directly in the study of conversation.

Equating information states with common grounds gives the update of an information state special logical properties. In update semantics, one of the ways to define logical consequence is by quantifying over information states σ: $\varphi_1, \ldots, \varphi_n \models \psi \Leftrightarrow \forall\sigma(\sigma \models \varphi_1, \ldots, \varphi_n \Rightarrow \sigma \models \psi)$. This cannot be maintained if σ ranges over common grounds rather than standard information states. Also, updating an information state to obtain another common ground is different from plain updating. The first half of this paper is concerned with common grounds as information states and their logic.

A distinction that is important and feasible is that between logical and pragmatic update operations. Logical updates correspond to what we are used to in logic and can be formally defined here as those operations that are eliminative and distributive over the information states that they update. They coincide with the operations that can be characterised by a Tarskian truth definition. The pragmatic ones (the speech acts and presupposition) can be defined as the ones that are not: they require properties of the information state as a whole in their definition. They typically also give rise to partiality. On this view, the presupposition operator (contra Beaver 1992) is a typical pragmatic operator. The second part of this paper studies some more of these operations. We end by an attempt to show that *might* is really a logical operator (contra Veltman 1996).

2 A basic update system

About the simplest possible update system is the one given by a language of propositional logic taking information states to be sets of models for that language.

We take all sets of models to be information states. The definition of update is given in (1). $[A]$ is the function from information states to information states, and we write $\sigma[A]$ for the result of applying $[A]$ to an information state σ. Updates are defined over this system by putting the update of an information state σ to be the intersection of the set of models in σ that satisfy φ.

(1) $\sigma[\varphi] = \{i \in \sigma : i \models \varphi\}$

We define $\sigma \models \varphi$ as an abbreviation of $\sigma[\varphi] = \sigma$, but we could equally well define it to be: $\forall i \in \sigma \; i \models \varphi$.

To this system we can add belief operators B. We assume a classical modal treatment: an operator B corresponds to an accessibility relation R_B between extended propositional models. The set $B_i = \{j : iR_Bj\}$ is the set of worlds that are accessible for B from a world i. This associates a set to every operator in every world. These sets can be thought of as the information state that B associates with i.

Kripke models could be used here, but instead I will follow Gerbrandy and Groeneveld 1996 in thinking of the elements of the information states as possibilities. A possibility is function that maps propositional letters to truth values and the belief operators to sets of possibilities. Using Aczel's non-well-founded set theory (Aczel 1988), we can show that possibilities exist. The main advantage is that we will have an easier time when discussing CG-updates later on.

We assume at least the system K for our belief operators. Assuming introspectivity ($B\varphi \rightarrow BB\varphi$) would not be problematic but adding reflexivity ($B\varphi \rightarrow \varphi$) would create problems for our common grounds[1]: we could no longer agree to disagree.

[1] Some material in the common ground is known, other material is only believed. This holds in particular for the material that we acquire in the course of communication

There are two different updating rules for the belief operator in the literature. One is due to Kamp[2], another due to Stalnaker[3]. The Kamp definition starts from the idea that the belief subject may have any kind of information. Some possibilities i will have $B_i \models \varphi$, others will not. So an update can be given by eliminating the possibilities where the subject does not believe the proposition attributed to her. This restricts the information state to possibilities in which the proposition holds in the belief state of the subject. The definition is given in (2).

(2) $\sigma[B\varphi] = \{i \in \sigma : B_i \models \varphi\}$

Stalnaker's way is to collect the belief information states in the different possibilities and collect them in one single information state by set union. This gives a single information state (what the subject believes in the information state) which is then updated by the proposition the subject is asserted to believe. We then check whether a possibility assigns to B an information state that has as least as much information as the information state that results from the update. If not, the possibility is eliminated.

(3) $\sigma[B\varphi] = \{i \in \sigma : B_i \subseteq (\bigcup_{i \in \sigma} B_i)[\varphi]\}$

For our basic system, both definitions coincide. For a distributive and eliminative system: $B_i \models \varphi$ iff $B_i \subseteq (\bigcup_{j \in \sigma} B_j)[\varphi]$. This does not always hold: Beaver (p.c.) shows that they diverge on Veltman's *might*-operator.

In the sequel, we will freely use both definitions[4].

3 Common Grounds

We can now embark on a discussion of the common ground. With a common ground, there is the set of participants whose common ground it is. We can equate these with a set of belief operators P in some set CGP (common ground partners). It seems reasonable to ask that CGP is a finite non-empty set of belief operators. The case that CGP has only one partner is special. For a common ground between P and himself, $\sigma \models \varphi$ iff $\sigma \models P\varphi$. They are the information states in which P's beliefs coincide with the available information: it is introspective belief.

The case of two participants seems representative of the case of more than 1 participant and we will sometimes assume that $CGP = \{S, H\}$ to facilitate discussion. A basic intuition

[2] As reported by Heim 1992

[3] Stalnaker 1988

[4] Both definitions continue to be the same when we generalise to a version of FOL. A discussion falls outside the scope of this paper.

is that the common ground contains the information that each participant shares with the other. But this is not sufficient as parties may agree with each other in certain respects without knowing so. When this happens, the material should not be in the common ground: the parties share it but they are not aware that they do so and therefore, they cannot draw on this material in their consideration of collaborative actions and communication with each other. We must strengthen our definition to read: the common ground contains all that information about which all parties, according to the common ground, agree. This is circular, but we can still employ it to single out among the information states those states that are common grounds. Let $P_1, \ldots, P_n$ be the belief operators of the participants. Then (4) is a necessary condition under which an information state can be a common ground.

(4) $\sigma \models \varphi \Leftrightarrow \sigma \models P_1\varphi \wedge \ldots \wedge P_n\varphi$

One of the things that we can prove in general is that the state of no information 1 and the state of inconsistent information 0 fulfill this condition. For 0, notice that $0 \models \varphi$ for any φ. Notice that for 1, $1 \models \varphi$ only if φ is a tautology. But $P\varphi$ is then also a tautology and on 1, $P\varphi$ only holds if φ is tautology.

I want to introduce another idea here, which may be more controversial. Participants in a common ground know that they are dealing with a common ground. That is they believe that whatever they believe to be common ground between them is the case according to the common ground. So the common ground comes with the pretense that what is believed in common is common ground, i.e. true according to the common ground. This makes it plausible to add another condition in a definition of common ground: whatever is shared holds. Since this is a shared belief, we want it to hold in the common ground. This leads to the definition of a common ground in (5).

(5) An information state σ is a common ground if and only if
$\sigma \models P_1\varphi \wedge \ldots \wedge P_n\varphi \rightarrow \varphi$ and
$\sigma \models \varphi \Leftrightarrow \sigma \models P_1\varphi \wedge \ldots \wedge P_n\varphi$

Updates of a common ground will not in general bring us from one common ground to the next, because the conditions may cease to hold.

In the sequel we will use $\Box\varphi$ as an abbreviation of $P_1\varphi \wedge \ldots \wedge P_n\varphi$.

Notice that what a participant believes in the common ground cannot be less than what the common ground itself contains as information. But it could well be that in some or all of the possibilities of the common ground the participant believes more. Think of the case that a participant has expressed a belief that has not been accepted by the others. This leads to a structural condition on common grounds σ.

(6) $P_i \subseteq \sigma$ for P a participant and $i \in \sigma$

(This follows from the demand: $\sigma \models \varphi \Rightarrow \sigma \models P\varphi$ and the assumption that CGs are uniquely determined by their theories.)

A second fact of this kind is the fact that all possibilities in a common ground must be allowed for by at least one participant. That is, the equation (7) holds for common grounds σ.

(7) $\sigma = \bigcup_{P \in CGP} \bigcup_{i \in \sigma} P_i$

One side of the equation follows from (6) . The other follows from lemma (8),

(8) **Lemma** $\sigma = \bigcup_{i \in \sigma} \Box_i$ where $\Box_i = \bigcup_{P \in A} P_i$

a lemma for which we also have to make the extra assumption that our information states are uniquely determined by their theories.

Our lemma then follows from lemma (9).

(9) **Lemma** $\sigma \models \varphi \Leftrightarrow \bigcup_{i \in \sigma} \Box_i \models \varphi$

Proof Let σ be a common ground and assume that $\sigma \models \varphi$. By the definition of common grounds, this is equivalent to $\sigma \models \Box\varphi$. By distributivity, this is the same as demanding that $\Box_i \models \varphi$ for each $i \in \sigma$. By a second application of distributivity this is the same as demanding that for every $i \in \sigma$ and for every $j \in \Box_i\{j\} \models \varphi$. But that is equivalent by distributivity to $\bigcup_{i \in \sigma} \Box_i \models \varphi$.

So the shared beliefs of the participants are subsets of the common ground and also form a cover of the common ground.

These lemmas suggest a direct semantic definition of the common ground. Let R be the union of the accessibility relations. :σ is a common ground iff $\{i : \exists j \in \sigma : jRi\} = \sigma$, and $R \cap (\sigma \times \sigma)$ is reflexive[5].

4 Updating the Common Ground

The problem that we have to face now is adding information to an information state that is a common ground in such a way that we end up with a common ground again. This problem is a variation of what has always seemed problematic about common grounds. If

[5] Thanks go to Gerd Jaeger for suggesting this definition

we add φ we have to add $P\varphi$ for each participant P as well. And if we have done so we must do this again for the new statements as well. And so forth ad infinitum.

In the same way, we must take care when we add a statement of the form $P\varphi$. Not only do we have to take care that this new statement gets added as beliefs of all the participants, but we also have to take care that it is not suddenly the case that a new bit of common ground has emerged as it can already be the case that all the other participants agreed about φ.

The problem faces us with almost every speech act. If we have only a speaker S and a hearer H, an assertion is only proper when it is in the common ground that the hearer does not believe the content of the assertion. It is reasonable to demand that the hearer does not believe the negation of the content, it should be common ground (or be accommodatable) that the hearer does not have the opinion that φ. If all goes well, after the assertion, it should be the case that the hearer now believes that φ. The extra evidence for the content of the assertion that has made her change her mind was the fact of the assertion. But this means that addition of information by communication is not a monotonic process: we must get rid of information and replace it by new information.

Let us look at this in some more detail. A successful assertion of p can be described as a transition from a common ground σ to an information state τ such that (10).

(10) $\tau \models p$
$\tau \models \Box p$

Things go wrong if we describe the speaker as making the assertion because of her assessment of the common ground: she must take it to be the case that it is not the case that the hearer believes that p, i.e. she must believe $\neg Hp$. So whether the assertion is successful or not, it is evidence that the speaker believes p and believes $\neg Hp$. If this is so, the assent of the hearer (leading to the desired result of the assertion) has to override the conflicting determination of the speaker's assessment of the hearer's attitude with respect to p.

This is not a mistake. What goes on in communication is a change in the world: first the hearer has no evidence for p, now she has. First there was no reason for H to open the window, now the request has provided a reason. First, S was under no obligation to do X, the promise has changed this.

I am not dealing with corrections, only with assertions where the speaker is adding facts consistent with the common ground and with the expressed beliefs of the hearer. It is conceivable that the hearer believes the negation of p but that the speaker is trying by her assertion and maybe by later argument to get her to change her mind. The treatment of such corrections is however difficult within the current setting. A reasonable first step is to find a way of updating that can deal with the problem of conflict-free updates. What remains open for conflict, is the road of keeping records of earlier information states. The common ground however does not seem to offer new ways of dealing with the problems of belief revision.

Let σ be given. Let p be the content of the assertion and consider $\sigma[p]$. $\sigma[p]$ will no longer be a common ground. Let us assume that σ has a p-possibility i in which the hearer does not believe that p. That means that H_i is partly outside $\sigma[p]$, as $\sigma[p]$ only contains p-possibility and H_i must have at least one possibility that is not a p-possibility. i will survive the update with p and thereby keep $\sigma[p]$ from being a common ground.

The intuition behind the following operation of restriction is the following. We want to change the possibilities in the information state that have the offending property by changing the possibilities to which the partners have access. We start by looking at a more general case. Consider the following operation of *restriction* of one information state by another given a fixed CGP.

(11) $\sigma^\tau = \{i^\tau : i \in \sigma\}$
$i^\tau(x) = i(x)$ for $x \notin CGP$
$i^\tau(x) = (i(x) \cap \tau)^\tau$ for $x \in CGP$

The operation is not recursive in set theory, but it is allright under Aczel's AFA. What should however be clear that the definition of $\sigma^\tau \models \varphi$ is recursive in the definitions of $\sigma \models \varphi$ and $\tau \models \varphi$. By our earlier assumption, it follows that we can think of σ^τ as an information state.

The operation limits the extension of the beliefs of participants in σ to the information state τ. We will use this operation to model the following situation: We update a common ground σ with some new information φ and then restrict the new information state by itself. This happens to be a common ground.

The self-application σ^σ can be written as an operation * mapping information states to information states.

So $\sigma[\varphi]* = \sigma[\varphi]^{\sigma[\varphi]}$. We can prove the following theorem (12) which is slightly more general than we require.

(12) **Theorem** If $\sigma \models \Box\varphi \to \varphi$ for all formulas φ, then $\sigma*$ is a common ground.

Proof.
We use τ for $\sigma*$ and let $\Box_j = P_{1_j} \cup \ldots \cup P_{n_j}$ and $\Box_\tau = \bigcup_{i \in \tau} \Box_i$.
We first show that $j \in \Box_j$ for $j \in \tau$.
Let $j \in \tau$. Then there is an $i \in \sigma$ such that $j = i^\sigma$. Because $i \in \Box_i$ (by $\Box\varphi \to \varphi$) and $i \in \sigma$, we have $j \in \Box_j$. From this it follows immediately that $j \in \Box_\tau$ and so that $\tau \subseteq \Box_\tau$. It also follows that $\tau \models \Box\varphi \to \varphi$.
The construction on the other hand guarantees that $\Box_\tau \subseteq \tau$. Combining, we have $\tau = \Box_\tau$ and it follows that $\tau \models \varphi$ iff $\tau \models \Box\varphi$.

In particular, it follows from the theorem that $\sigma[\varphi]*$ is a common ground if σ is a common ground.

What do we know about our new common ground $\sigma[\varphi]*$? First of all, if a formula holds on σ and it does not contain any occurrence of P, it will continue to hold on $\sigma[\varphi]*$. Second, for such formulas it also holds that they will continue to hold when prefixed with a P: the participants' belief sets become smaller. What can stop holding are negations of $P\varphi$ and this is as it should be: the information of the participants has increased and they believe more than they used to. Also, if there is a possibility in which $P\varphi$ holds, with φ free of P's, $\sigma[\varphi]*$ will also have such a possibility.

(13) **Lemma** Let φ range over formulas in which there is no operator P, with P taken from CGP. Then:
1. $\sigma \models \varphi$ iff $\sigma^\tau \models \varphi$
2. If $\sigma \models P\varphi$ then $\sigma^\tau \models P\varphi$
3. If $\sigma \not\models \neg P\varphi$ then $\sigma^\tau \not\models \neg P\varphi$(Here and in (2) P may be a sequence of operators from CGP.
4. If $\tau \models \varphi$ then $\sigma^\tau \models \Box\varphi$

From (13) it follows that $\sigma[\varphi]* \models \varphi$ whenever φ is equivalent to a positive formula over the operators in CGP. How about the other formulas? A curious example is Moore's paradox:[6] on $\sigma[p \wedge \neg Pp]*$ the formula $p \wedge \neg Pp$ does not hold but Pp. Another example of an update that leads to its negation is given by Gerbrandy and Groeneveld using the Conway-paradox. It is the $*$-operation that is to blame here: it extends the knowledge of the conversationalists. This shows that common ground updating does not obey the principle: $\sigma[\varphi] \models \varphi$ and so can easily fail as a characterisation of logical operators, unlike the basic update system we considered before.

[6] $\sigma[p] * [\neg Pp]*$ is the inconsistent information state

5 Computation

In abstracto, it is hard to deal with common grounds, in practice much easier, at least if we restrict ourselves to what is needed for a theory of communication.

In communication, a common ground can always be thought of as being a basis which is closed off under the schemes we have been discussing. The basis is the set of those facts which cannot be derived by the schemes from other facts. For the analysis of real conversations, it would (as judged by one of the participants) consist of what she has in common with the other party in knowledge of the language used and in world knowledge, what aspects of the speech situation are shared and finally of her own commitments and those of the other participants.

In practice, world knowledge and knowledge of language can be reduced to that part that has been actively used in the communication at hand. Use of a word or expression indicates knowledge of language, inferences indicate the acceptance of certain world knowledge.

Our characterisation of the common ground gives us two schemes that can be almost directly used. For the second scheme, it is just a matter of adding the formula scheme $\Box\varphi \to \varphi$ to the basis. This will also take care of one half of the first scheme: the part that says that if $\sigma \models \Box\varphi$ then $\sigma \models \varphi$.

For the rest we need closure under the rule: $T \vdash \varphi$ then $T \vdash \Box\varphi$, where T is the extended basis.

To sum up, if we assume that T is a CG-basis then $T \vdash_{CG} \varphi$ iff $\varphi \in S$ where S is the smallest set containing T, and all instances of $\Box\varphi \to \varphi$ which is closed under K and the rule $S \vdash_K \varphi \Rightarrow \Box\varphi \in S$.

But what is T? It is reasonable to allow common grounds to start from somewhere: general knowledge of the kind that is described as knowledge of language and world knowledge is one ingredient. The other components can be taken as consisting of two elements: a characterisation of the speech situation and the commitments of speakers and hearers. The last element directly corresponds to the commitment slates due to Hamblin 1971 and put to action by Van Leusen in the context of corrections. It appears therefore that there is little difference between our CG-updating and maintaining commitment slates, rather the two views of maintaining conversational information are complementary. Commitment slates are a practical answer to how to maintain a common ground as a conversation unfolds, CG-updating supplies an answer to the question what the meaning of the commitment slate is and what consequences can be drawn from a given commitment slate.

Commitment slate updating corresponds to CG updating. If an assertion is made we can add to the commitments: $S\varphi$, if the assertion is accepted we add φ. If the hearer rejects it, we add $H\neg\varphi$ and so on. We will study this process more closely in the next section. Things become common ground, because both speaker and hearer believe it and there is no reason to add a special section in a commitment slate which maintains common beliefs.

There is also no reason for limiting oneself to formulas of a particular logical complexity.

6 Applications

Our basic system is both eliminative and distributive. That means it is not necessary to treat it as an update semantics at all[7]. This changes as soon as we switch to the operations on common grounds in which we are really interested: speech acts, presupposition resolution, querying and epistemic modalities.

The notion of information can be understood as a test[8]. We imagine a subject who we tell that she is placed in a possible state of affairs and we want her to tell us whether the possible state of affairs is the actual world or not. There is no limit on the amount of investigation of the alternative the subject can engage in. Now the criterion is: can a conflict between (in principle discoverable) facts in the state of affairs and the information of the subject be constructed. If there is such a conflict, the subject will conclude that no, this is not the actual world, otherwise she will not be able to decide whether it is or not.

What happens of course in the test is that we keep the information constant: this is the resource for the subject to carry out the test. Now it seems that information about information is typically what speech acts are involved with, and it seems right that we separate this off from information of the kind that gives a criterion for deciding that a possible state of affairs is not the actual world. An assertion is an indication that I have certain information, a query an indication that the speaker wants to have such and such information, a presupposition an indication that the speaker takes such and such information for granted, etc.

What we attempt below is to use our framework as a means for defining the basic moves in conversation. I will call such moves speech acts. This is appropriate, as they share important characteristics with actions. A speech act is essentially a way to change the common ground in a controlled way. It can take effect only under certain circumstances (the preconditions) and has both a basic effect and intended effects. The basic effect is always reached, the intended effect depends on further speech acts of the hearer. A question can only be put under certain conditions, e.g. that the speaker knows the answer cannot be common ground. It has as direct effect to make it common ground that the speaker does not know the answer. But the intended effect is that the answer will become common ground, a goal that is only reachable through the participation of the hearer.

[7] It has been shown that such semantics can always be dealt with in a static way
[8] I am indebted here to Haas-Spohn, see her 1995

6.1 Assertion

The most basic case is the assertion. For a proper assertion (not a correction or a self-correction or a reiteration) it must be the case that the common ground does not deny its content, that the speaker is not known to deny its content and that the hearer is not known to deny its content. In all these cases it is a correction of some kind. It should also not be the case that the content is known, known by the speaker or known to the hearer. (Here the fact that the speaker (or the hearer) does not know it or its negation entails that it is not CG.)

In our setting each of these means that there is a possibility in the CG in which the content is not true and this is not a fact that can be inspected by looking at one possibility only. So the fact that an assertion is proper given a CG (otherwise it would be undefined) cannot be seen as a distributive and eliminative update. Much the same holds for presupposition and the epistemic modalities.

It turns out that for the definedness of special updates, corresponding to speech acts, it is necessary to look at what information is not contained in the information state. This is not a distributive test, as non-satisfaction, in a distributive eliminative update semantics comes down to the existence of a carrier that does not satisfy the proposition in question.

A good case is assertion. I start from a conception of assertion where the assertion is carrying out one of the useful functions of communication: to supply information that one could in principle acquire by one's own observation but which one has not observed oneself. The asserter is here ideally the end of a chain going back to an original observation of the asserter or of someone that has transmitted, directly or indirectly, the information to the asserter by communication. The fact that someone asserts something then has a comparable status to observation itself: it is evidence for the truth of the content of the assertion. If we take assertions to be an attempt on the part of the speaker to change the common ground to contain some information it did not previously contain, by means of the evidence constituted by the speaker's assertion, we get the four demands in (14) on what the common ground should be like if the assertion is to be successful[9].

(14)
$$\sigma \not\models H\varphi$$
$$\sigma \not\models H\neg\varphi$$
$$\sigma \not\models S\varphi$$
$$\sigma \not\models \neg S\varphi$$

As we are dealing with a common ground these four conditions entail the Stalnaker conditions given in (15).

(15)
$$\sigma \not\models \varphi$$
$$\sigma \not\models \neg\varphi$$

[9] Jelle Gerbrandy noted two important problems in the original treatment

But the Stalnaker conditions are weaker: they entail that two of the four conditions hold. The common ground can e.g. be as in (16). It follows from these two demands (and σ not being the absurd state) that the Stalnaker conditions hold. On the basis of this example, it seems fair to conclude that the Stalnaker conditions are too weak.

(16) $\sigma \models H\neg\varphi$
$\sigma \models S\varphi$

The four conditions can be justified as follows. If the hearer would already (be known to) know the content of the assertion, the assertion could not change the common ground in the sense of adding the content of the assertion to it. If it is possible to utter φ at all when this condition applies, we would be dealing with the assent of the speaker to a previous assertion of the hearer. This is not an assertion.

If the hearer would be known to believe the assertion is false, we are ready for conflict. This is certainly possible, but it is useful to distinguish this case from proper assertions and reserve the word correction for that. The speaker can certainly not expect by his utterance of φ alone to change the common ground in the desired direction.

If the speaker is known to believe the content of the assertion already, it seems again that the assertion by itself will not be sufficient to change the common ground. The speaker repeats his previous statement and obviously the hearer did not believe him before.

Finally, if the speaker is known not to know the assertion, it is unclear by what means he hopes to change the common ground. The pretense associated with the use of an assertion is that the speaker has acceptable evidence for his belief in the truth of the assertion. If it is known he does not know it the fact that he asserts it will not be evidence for the truth of the assertion.

The assertion of φ will minimally indicate that the speaker believes that φ. So after one step, we reach (17).

(17) $\sigma[S\varphi]*$

The choice is now to the hearer: he can assent, express his disbelief or express his doubt. This brings us to the states in (18) respectively.

(18) $\sigma[S\varphi] * [\varphi]*$
$\sigma[S\varphi] * [H\neg\varphi]*$
$\sigma[S\varphi] * [\neg H\varphi]*$

In the first case the speaker reaches his goal.

In the other cases, there is now clarity about the hearer's opinion about the question whether φ.

Perhaps, I should say something about the other cases: the pseudoassertions arising by the failure of one of the conditions. First of all, if the speaker is assenting to the hearer, the assertion is automatically successful (unless the hearer has a general reason for distrusting the speaker) as the assertion at least is evidence for the speaker's belief in the proposition. The addition of the speaker's belief makes the content a part of the common ground: the evidence for φ in the speaker's assertion is not the reason for it becoming common ground.

When the speaker utters φ against the opposite view of the hearer, the strategy of the speaker must be different from just adducing evidence for φ by asserting it. The speaker can count on his position of authority, on the force of the arguments he is going to bring in later on, but perhaps his goal is also a more modest one: to bring about doubt or to bring about a deadlock in the communication.

Second, it is possible to reiterate what one has said before, and it even appears that this can carry out a useful function in the flow of communication. We can get back to earlier phases in the communication in this way or we can identify objects that have been referred to before. But it is clear as well that we do not adduce further evidence for the content by such a reiteration.

Last, it is also possible to correct oneself. This requires further justification: why one was wrong before and now is right. A normal isolated assertion will not lead to the goal.

6.2 Other Speech Acts

If one considers other speech acts like the question, the request and the promise, things are not very different.

Let us assume (for convenience) that we are dealing with a *yes-no*-question. The question is correct if the conditions in (19) are satisfied.

(19)
1. $\sigma \not\models S\varphi \vee S\neg\varphi$
2. $\sigma \not\models H\varphi$
3. $\sigma \not\models H\neg\varphi$
4. $\sigma \not\models \neg(H\varphi \vee H\neg\varphi)$

If the speaker would be known to know the answer to the question, his purpose of eliciting the answer from the hearer would be defeated (It would be a rhetorical question). If the hearer is known to know a particular answer, similarly the purpose of eliciting the information is not achievable by the question as it has already been reached before the question is asked. Last, if it is known that the hearer does not know the answer, there is again no purpose in asking the question. Together the four conditions entail (20).

(20) 5. $\sigma \not\models S\varphi$
6. $\sigma \not\models S\neg\varphi$
7. $\sigma \not\models \varphi$
8. $\sigma \not\models \neg\varphi$

In addition to the four conditions, we should be able to assume that it is consistent to assume that the speaker wants to have information from the hearer. I will stay clear from questions of desire[10], but just offer the extra demand in (21).

(21) $\sigma \not\models \neg want(s, S\varphi \vee S\neg\varphi)$

Putting the question the speaker changes the common ground to (22).

(22) $\sigma[want(s, S\varphi \vee S\neg\varphi])]*$

The hearer can answer *yes* or *no* or can deny to know the answer or can refuse to answer, changing the common ground to respectively (23).

(23) $\sigma[want(s, S\varphi \vee S\neg\varphi])] * [H\varphi]*$
$\sigma[want(s, S\varphi \vee S\neg\varphi])] * [H\neg\varphi]*$
$\sigma[want(s, S\varphi \vee S\neg\varphi])] * [\neg H\varphi \wedge \neg H\neg\varphi]*$
$\sigma[want(s, S\varphi \vee S\neg\varphi])] * [\neg wants(h, S\varphi \vee S\neg\varphi)*$

Assents and denials can then further bring φ or its negation in the common ground. (In case the hearer gives a positive or negative answer, the speaker's desire is fulfilled and can be eliminated. One way of achieving this is by preference semantics).

An interesting observation about questions is that their preconditions are the ones that make any answer to it a proper assertion. This supports the view that all assertions must be seen as answers to (possibly hidden) questions.

It is possible to steer completely free from the moral dimension. The common ground is an assumed object for the speaker and the hearer which is manipulated by them both according to what they think is happening. Lies are occasions where the speaker manages to insert things in the common ground he knows are false, false promises occasions where a promise is made without the intent to carry it out. This may be immoral, but it changes little as to the communication itself: as always, we build a faithful picture of what has happened between speaker and hearer and keep a list of what they want, plan or believe.

[10] It seems that it is possible to maintain a set of shared desires in the common ground.

6.3 Might

One of the motivations for developing this account of common ground updating is a dissatisfaction with the semantics of the might-operator proposed by Veltman. The semantics that Veltman proposes is (24).

(24) $\sigma[might\varphi] = \sigma$ if $\sigma[\varphi] \neq 0$ and otherwise 0.

This semantics does not connect well with the standard idea that assertive updates give information, which in our context is equivalent to them eliminating at least some possibilities. Of course updates with *might* φ may eliminate all possibilities, but this would be too much: if we move to the inconsistent information state we have lost everything and it may be right to assume —with Stalnaker— that one of the principles guiding our interpretation system is to avoid landing in the absurd information state. So in both cases, there is conflict with the Stalnaker conditions.

We may of course question whether an utterance of *might* φ is indeed an assertion. This may be fruitful but runs counter to the intuition that indeed utterances of *might* φ normally supply extra information, an intuition which is the basis for wanting to classify *might* φ as an assertion.

Let us however proceed from the opposite view. We will try to analyse *might* as a speech act operator and then show that its effect can be captured by assuming it is a logical operator.

Assume then that utterances of *might* φ are no assertions but speech acts of a kind of their own and let us try to analyse this new class of speech acts in the way we did before.

It seems reasonable to assume the preconditions in (25).

(25) $\sigma \not\models \neg S \neg \varphi$
$\sigma \not\models \neg H \neg \varphi$
$\sigma \not\models S \neg \varphi$
$\sigma \not\models H \neg \varphi$

The first two make the utterance have a purpose, the second two ensure that we steer clear of conflicts.

From these it follows that also the conditions in (26) hold.

(26) $\sigma \not\models \varphi$
$\sigma \not\models \neg \varphi$
$\sigma \not\models S \varphi$
$\sigma \not\models H \varphi$

Does *might* φ have a contribution? It would appear that the least that is required is the speaker does not think that $\neg\varphi$ is the case. So the contribution of the speaker's speech act in his utterance of *might* φ is the common ground (27).

(27) $\sigma[\neg S\neg\varphi]*$

The hearer has a choice of reactions: assent, denial or doubt.

Assent would be the further change to (28). (There is not much point in assenting to the speaker's disbelief).

(28) $\sigma[\neg S\neg\varphi] * [\neg H\neg\varphi]*$

Denial would have to take the form of an assertion (!) of $\neg\varphi$ and doubt would be the impossibility for the hearer to decide between his knowing or not knowing that $\neg\varphi$. This is $\neg HH\neg\varphi \wedge \neg H\neg H\neg\varphi$ which admittedly is a somewhat sophisticated attitude to have towards a proposition. (A: Maybe John is home. B: I don't know). So far so good. Notice that *might* φ gives new information. If accepted by the hearer, it makes ignorance of $\neg\varphi$ common ground.

Suppose the above is correct. We may then represent *might* φ as $\neg S\neg\varphi \wedge \neg H\neg\varphi$ which we can abbreviate as $\Diamond\varphi$ (notice that this is not the diamond belonging to our earlier necessity operator or the diamond defined by $\neg B\neg$).

This will give us the following preconditions (29) on an utterance of *might* φ as instances of the assertion precondition discussed before.

(29)
$$\begin{array}{l} \sigma \not\models S\Diamond\varphi \\ \sigma \not\models S\neg\Diamond\varphi \\ \sigma \not\models H\Diamond\varphi \\ \sigma \not\models H\neg\Diamond\varphi \end{array}$$

This entails two of our previous conditions, i.e. (30).

(30)
$$\begin{array}{l} \sigma \not\models S\neg\varphi \\ \sigma \not\models H\neg\varphi \end{array}$$

but not the other two we had before:

(31)
$$\begin{array}{l} \sigma \not\models \neg S\neg\varphi \\ \sigma \not\models \neg H\neg\varphi \end{array}$$

Instead we only get the weaker condition (32).

(32) $\sigma \not\models \neg S \neg \varphi$ or $\sigma \not\models \neg H \neg \varphi$

It would appear though that this condition is more correct than the earlier one. Suppose it is common ground that S does not know that not φ. Our precondition then entails that it is not common ground that H does not know this. H's assent would add something to the common ground.

(33) S: I do not know that φ is false. It might be that φ.
H: Yes, it might.

Inversely, suppose that it is common ground that the hearer does not know that φ is false, e.g. because the hearer has asserted φ before. Then the condition boils down to it not being common ground that the speaker does not know that $\neg\varphi$. This seems very natural, witness (34)

(34) H: John is ill.
S: He might.

So, it seems clear that our earlier conditions are too strong and that the current ones are better.

Also, the contribution changes slightly (and I think unimportantly):

(35) $\sigma[S\Diamond\varphi]*$

The hearer can assent by making the common ground into (36).

(36) $\sigma[S\Diamond\varphi] * [H\Diamond\varphi]*$

Denial would indeed be the negation of φ and the curious declining of the speaker's proposal would be equivalent to asserting $\neg\Diamond\varphi$, which given the fact that $\neg S\neg\varphi$ has been established comes out as the sophisticated doubt about the hearer's disbelief that we found before.

I conclude that $\Diamond\varphi$ is as good an approach to *might* in the current context as the separate speech act theory. It moreover makes *might* a logical operation with a distributive and eliminative update.

One can wonder however whether we have captured the meaning of might, and, indeed, I am not convinced. Suppose John is a BSE expert to whom we ask: Can the consumption of cheese lead to BSE? We of course do not have a clue, that is one of the reasons we ask this to John. John now says: it might. John seems to speak not so much on behalf of us, the conversational partners, but on behalf of his professional group: The BSE experts have not been able to rule this out.

Consider further the embedded use of *might* in e.g. *John thinks it might rain.* It seems obvious that neither the speaker's opinion nor the hearer's opinion as to whether it rains has any bearing on the truth of this attribution. In (37), there are two examples.

(37) John is home but Bill thinks he might be at work.
John is at work but Bill only thinks that he might be home.

Both doubts point in the direction of conceiving of *might* as an epistemic operator which claims of a certain group of people that they do not have the information to rule out the complement. The group of people would be determined by the context, much like a pronoun. The group must obey one constraint: the speaker or thinker must be inside it. In a conversation, when *might* does not appear in a propositional attitude context as generated by verbs like believe, know or say (these verbs would change the identification of the group, as they may change the identity of the speaker or thinker) a very natural resolution of the group parameter is the group of the conversational partners. So the analysis we provided is only a special case.

It can also be shown that the "stability facts" from Veltman's paper around *might* are undisturbed. We can have sequences

(38) *might* φ. φ
might $\varphi.\neg\varphi$
might φ. *might* $\neg\varphi$

but not sequences like

(39) φ. $\neg\varphi$.
φ. *might* $\neg\varphi$
φ. *might* φ.

All of these acceptabilities and inacceptabilities can be explained from the assertion preconditions.

To sum up, we have presented a theory of might which makes it into a logical operator which exhibits both distributivity and eliminativity. We concur with Veltman in his assumption that (normally) might sentences do not affect the factual basis of the common ground and in the contention that there are certain stability facts around might. We do not think however that these observations lead to an analysis of simple might-sentences which makes them into non-assertions or even turns *might* into a pragmatical operator.

7 Conclusion

This paper was written in response to an observation and a worry by David Beaver (p.c.). The observation was that a sentence *might*φ (with the Veltman semantics) is a counterexample to the equivalence between the Kamp and the Stalnaker update rule for belief. Indeed the Stalnaker rule leads to incorrect results. The worry was that common ground updating might well be inconsistent. The observation is devastating for any theory of presupposition resolution and accommodation in update semantics which wants to treat belief contexts, as using the Kamp rule would make the choice between resolution and accommodation or the choice between different accommodations dependent on individual possibilities, whereas these choices determine the global interpretation of the sentence. For a proper treatment, we need the Stalnaker rule. The equivalence can only be maintained, if we find no operations in the sentence which are either non-distributive or non-eliminative. A presupposition operator as proposed by Beaver Beaver 1992 would be the other candidate that I know of. And I just indicated that it coexists badly with the Kamp rule.

This work needs follow-up in three directions. The first is an obvious one: first order logic, which will also allow more questions. The second direction, is to find out more about corrections. We can now only state correctness conditions for corrections, but we want to be able to actively retract material and, importantly, to guarantee the continuation of the common ground. If the effect of retraction is not public, the new common ground is not public and therefore not a common ground. The third direction is to incorporate more than just facts in a common ground. We can have joint public goals and obligations and there are speech acts involving goals and obligations. It remains to be seen whether these can be incorporated in our model.

References

Aczel, P.: 1988, *Non–Well–Founded Sets*, No. 14 in CSLI Lecture Notes, CSLI Publications, Stanford

Beaver, D. I.: 1992, The kinematics of presupposition, in *Proceedings of the Eighth Amsterdam Colloquium*, ILLC, Amsterdam

Gerbrandy, J. and Groeneveld, W.: 1996, *Reasoning about Information Change*, ILLC preprint LP-96-10

Haas-Spohn, U.: 1995, *Versteckte Indexikalität und subjektive Bedeutung*, studia grammatica 38, Akademie-Verlag, Berlin

Hamblin, C. L.: 1971, Mathematical models of dialogue, *Theoria* 37, 130–155

Heim, I.: 1992, Presupposition projection and the semantics of attitude verbs, *Journal of Semantics* 9, 183–221

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

Stalnaker, R.: 1978, Assertion, in P. Cole (ed.), *Pragmatics*, No. 9 in Semantics & Pragmatics, pp 315–32, Academic Press, New York

Stalnaker, R.: 1988, Belief attribution and context, in R. Grimm and D. Merrill (eds.), *Contents of Thought*, University of Arizona Press

Veltman, F.: 1996, Defaults in update semantics, *Journal of Philosophical Logic* 25