VIII. Theories of discourse semantics

36. Situation Semantics: From indexicality to metacommunicative interaction

1. Introduction
2. Desiderata for semantics
3. The Relational Theory of Meaning
4. Meaning, utterances, and dialogue
5. Closing remarks

Abstract

Situation Semantics emerged in the 1980s with an ambitious program of reform for semantics, both in the domain of semantic ontology and with regard to the integration of context in meaning. This article takes as its starting point the focus on utterance (as opposed to sentence) interpretation. The far reaching aims Barwise and Perry proposed for semantic theory are spelled out. Barwise and Perry’s Relational Theory of Meaning is described, in particular its emphasis on utterance situations and on the reification of information. The final part of the article explains how conceptual apparatus from situation semantics has ultimately come to play an important role in a highly challenging enterprise, modelling dialogue interaction, in particular metacommunicative interaction.

1. Introduction

Situation Semantics emerged in the 1980s with an ambitious program of reform for semantics, both in the domain of semantic ontology and with regard to the integration of context in meaning. In their 1983 book Situations and Attitudes (Barwise & Perry 1983), as well as a host of other publications around that time collected in Barwise (1989) and Perry (2000), Barwise and Perry argued for the preeminence of a situation-based ontology and took contexts of utterance to be situations, thereby offering the potential for a richer view of context than was available previously. For situation semantics and ontology, see article 35 (Ginzburg) Situation Semantics and NL ontology. This article takes as its starting point the focus on utterance (as opposed to sentence) interpretation. In section 2 I spell out the far reaching aims Barwise and Perry proposed for semantic theory. In section 3 I sketch Barwise and Perry’s Relational Theory of Meaning, in particular its emphasis on utterance situations and on the reification of information. I also point out some of the weaknesses of Barwise and Perry’s enterprise, particularly the approach to context. One of these weaknesses, in my view, is that the theory is quite powerful, but it was, largely, applied to dealing with traditional, sentence-level semantics. The final section of this article, section 4, explains how conceptual apparatus from situation semantics has ultimately come to play an important role in a highly challenging enterprise, modelling dialogue interaction.
2. Desiderata for semantics

Barwise and Perry’s starting point is model theoretic semantics, as developed in the classical Montague Semantics tradition (see e.g. Montague 1974; Dowty, Wall & Peters 1981; Gamut 1991 and article 33 (Zimmermann) Model-theoretic Semantics): a natural language is likened to a formal language (first order logic, intensional logic etc). On this approach, providing a semantics for such a language involves primarily assigning *interpretations* (or *denotations*) to the words of the language and rules that allow phrases to be interpreted in a compositional manner. This allows both the productivity of NL meaning and the potential for various kinds of ambiguity to be explicated. Contexts, on this view, are identified with *indices* – tuples consisting of a small and fixed number of dimensions, prototypically providing values for *speaker, addressee, time, location*. Interpretations of words/phrases are then all taken to be relative to contexts, thereby yielding two essential semantic entities: *characters/meanings* which involve abstracting away indices from *contents/interpretations*. These – supplemented by lexical meaning postulates – can be used to explicate logically valid inference.

Barwise and Perry view this picture of semantics as significantly too restrictive. The basic perspective they adopt is one in which linguistic understanding is assimilated to the extraction of information by resource bounded agents in their natural environment (inspired in part by the work of Gibson, e.g. Gibson 1979). This drives their emphasis on a number of unorthodox seeming fundamental desiderata for semantic theory, desiderata we will subsequently come to see find considerable resonance in the desiderata for a theory of meaning for conversational interaction.

The first class of desiderata are metatheoretical in nature and can be summed up as follows:

**Desideratum 1** *The priority of information:* *language has external significance,* as model theoretic semantics has always emphasized, *but, as cognitive scientists of various stripes emphasize,* it also has mental significance, yielding information about agents’ internal states. *What is needed is a way of capturing the commonality between the external and the mental,* the flow of information – the chain from fact to thought in one participant’s mind to utterance to thought in another participant’s mind, graphically exemplified in Fig. 36.1.

![Fig. 36.1: The Flow of Information. From Situations and Attitudes, p. 17](image-url)
An important component in fulfilling this desideratum, according to Barwise and Perry, is a theory by means of which external (and internal) reality can be represented – an ontology of some kind. This is what developed into situation theory and type theory with records (see article 35 (Ginzburg) *Situation Semantics and NL ontology*). A key ingredient in such a theory are some notion of *constraints*, a way of capturing necessary, natural, or conventional linkages between situations (e.g. smoke means fire, image being such and such means leg is broken etc.), along with a theory of how agents in a situation extract information using constraints. The other crucial component is the naturalization of linguistic meanings – their reduction to concepts from the physical world – in terms of *constraints*.

The other two pivotal desiderata put forward by Barwise and Perry are more directly aimed at repositioning the semantic fulcrum, from interpretation towards context.

**Desideratum 2** INFORMATION CONTENT is UNDERDETERMINED BY INTERPRETATION

We might provide news about the Argentinean elections using any of the following sentences in (1). All three sentences uttered in these circumstances intuitively have the same external significance – we would wish to identify their content and, on some accounts, their meaning as well. Nonetheless, different information can be acquired from each: for instance, (1b) allows one to infer that Kirchner is a woman, whereas Lavagna is a man.

\[
(1) \begin{align*}
\text{(a)} & \quad \text{Kirchner beat Lavagna}. \\
\text{(b)} & \quad \text{Señora Kirchner defeated Señor Lavagna}. \\
\text{(c)} & \quad \text{Cristina’s losing opponent was Lavagna}.
\end{align*}
\]

**Desideratum 3** LANGUAGE is an EFFICIENT MEDIUM

Barwise and Perry emphasize that the flip side of productivity gets less attention as a fundamental characteristic of NL: the possibility of reusing the same expression to say different things. Examples of the phenomena Barwise and Perry had in mind are in (2), which even in 2009 are tricky. By ‘tricky’ I don’t mean we lack a diagnosis, I mean there is no single formal and/or implemented semantic/pragmatic theory that picks them all off with ease, interfacing along the way with inter alia theories of gesture, gaze, and visual access.

\[
(2) \begin{align*}
\text{(a)} & \quad \text{A: I’m right, you’re wrong. B: I’m right, you’re wrong.} \\
\text{(b)} & \quad \text{I want you, you, and you to stand here and I want you, you, and you to stand here. (based on examples in Levinson 1983; Pollard & Sag 1994)} \\
\text{(c)} & \quad \text{A: John is irritating John no end. B: He can’t be annoying him so badly.} \\
\text{(d)} & \quad \text{In last week’s FoLLI dissertation prize meeting sadly the linguist voted for the linguist, whereas the logician voted for the logician. (based on an example in Cooper 1996)}
\end{align*}
\]

3. The Relational Theory of Meaning

At the heart of situation semantics is the relation theory of meaning. There are two fundamentally innovative aspects underlying this theory, which bear significant importance to current semantic theory in the wider sense:
(3) a. Meaning reification: the reification of meanings as entities on which humans reason (rather than as metatheoretical entities, as standard in logic).

b. Speech Events as Semantic Entities: recognition of speech events (incl speakers, addressees, the speech token) as fundamental semantic units; sentences are viewed as derivative: type-like entities that emerge from utterances, or, as Barwise and Perry put it, uniformities over utterances.

To get a feel for the theory, consider a simple example. (4b), taken to be the meaning of (4a), is a crude representation of an early version of the Relational Theory of Meaning: a (declarative) meaning relates all utterance events \( u \) in which there exists a speaker \( a \), addressee \( b \), spatiotemporal locations \( l, t \), referents \( j, m \) (for the names ‘Jacky’ and ‘Molly’ respectively) to described events \( e \) in which \( j \) bites \( m \) at \( t \). This relation is exemplified graphically in Fig. 36.2, which emphasizes the reality of the utterance situation.

I have purposely used quasi-Davidsonian notation (see article 34 (Maienborn) Event semantics) to indicate that the central insight there is independent of the various more and particularly less standard formalisms in which the Relational Theory of Meaning has been couched. As we will soon see, there are various ways which differ significantly to cash out the characterization of \( u, e \) and their interrelation.


b. \[
\{ u, e \mid \exists a,b,l,j,m,t[uttering(a, 'Jacky is biting Molly', u) \land \text{addressee}(u, b) \land \\
\text{In}(u, l) \land \text{referring}(a, j, 'Jacky') \land \text{Named}(j, 'Jacky') \land \text{referring}(a, m, 'Molly') \land \\
\text{Named}(m, 'Molly') \land \text{coincident}(l, t) \land \text{describing}(a, e) \land \text{bite}(e, j, m, t)] \}
\]

Fig. 36.2: The meaning of ‘Jacky is biting Molly’ as a relation between situations in which this construction is uttered and events in which a Jacky bites a Molly.

From Situations and Attitudes, p. 122

Of the two assumptions, Speech Events as Semantic Entities was introduced by Barwise and Perry in a stronger form than (3), graphically exemplified in Fig. 36.2 – not only do they make reference to speech event, but Barwise and Perry actually posit a compositional aspect to speech events:

(5) a. If \( \alpha \) is a phrase with sub-constituents \( X, Y \), then uttering \( (a, \alpha, u) \) entails the existence of two subevents of \( e_1, e_2 \) such that
VIII. Theories of discourse semantics

b. \( e_1 < e_2 \) (\( e_1 \) temporally precedes \( e_2 \))
c. uttering(\( a, X, u_1 \))
d. uttering(\( a, Y, u_2 \))

This formulation raises a variety of issues concerning syntax, presupposing essentially a strongly surfacely and linearized approach. For obvious reasons of space I cannot enter into these, but they constitute an important backdrop. Speech Events as Semantic Entities underlay a number of grammar fragments subsequent to Barwise & Perry (1983) (e.g. Gawron & Peters 1990; Cooper & Poesio 1994), but on the whole was not the focus of much interest until Poesio realized its significance for conversational processing, as we discuss in section 4.2. In contrast, issues concerning Meaning reification drove much research in the hey day of Situation Semantics. The relation exemplified in (4b) is certainly a relation with relevance to the semantics of utterances of ‘Jacky is biting Molly’: it relates events in which the speaker mouths a particular linguistic form while referring to a Jacky and a Molly with an event the speaker is describing in which that Jacky bit that Molly. Barwise and Perry view \textit{attunement} – the awareness of similarities between situations and of relationships that obtain between such similar situations – to the constraint in (4) as being what underlies our competence to use and understand such utterances. Nonetheless, there are two aspects which the formulation above abstracts away from: contextual parameter instantiation and truth evaluation. (4) does not make explicit the fact that understanding such an utterance involves finding appropriate referents for the two NP sub-utterances, as indeed in certain circumstances – e.g. for an overhearer who cannot see the speech participants or hears a recording – for the speaker and the time. In fact, in the original formulation of the Relational Theory of Meaning Barwise and Perry made a point of not packaging all of context in one event/situation, but distinguished three components of context: (a) \textit{the discourse situation}, comprising the public aspects of an utterance (including all the standard indexical parameters), (b) \textit{the speaker connections}, comprising information pertaining to a speaker’s referential intentions, and (c) \textit{resource situations}, events/situations distinct from the described situation, used to serve as referential/quantificational domains. Although the discourse situation/speaker connection dichotomy does not seem to have survived – examples such as (2b) illustrate the importance of speaker intention even with ‘simple indexicals’, the ultimate insight to be drawn here, it seems, is the unbounded nature of contextual dependence. Resource situations are one of the important contributions of situation semantics (see particularly Cooper 1996), and are further discussed in article 35 (Ginzburg) \textit{Situation Semantics and NL ontology}.

Returning to (4), the formulation of the Relational Theory of Meaning as a relation between contextual situations (the discourse situation, speaker connections, zero or more resource situations) and described situations, is problematic. It means that the latter cannot serve as the denotations of declarative utterances (since they are not truth bearers), nor does it generalize to non-declarative meaning. This reflects the fact that in \textit{Situations and Attitudes} Barwise and Perry attempt to stick to an avowedly “concrete” ontology, one which eschews abstract entities such as propositions, leading them into various foundational problems.

This stance was abandoned soon after – various notions of propositions emerged as situation theory developed. Hence, in works such as Gawron & Peters (1990), Cooper & Poesio (1994), works from a maturer version of situation semantics, (declarative)
sentential meanings came to be formulated as relating utterance situations – from whence values for contextual parameters would be drawn – and propositions; meanings for sub-sentential constituents would analogously relate an utterance situation for that constituent with an associated described object (referent [NP], property [VP] etc).

As an example of the Relational Theory of Meaning in a current formalism that fixes both problematic aspects discussed above, consider (6), which uses the formalism of Type Theory with Records (see Cooper 2006), discussed in more detail in article 35 (Ginzburg) Situation Semantics and NL ontology. (6a) corresponds to an utterance type (utterance type in the sense of sign as in constraint-based grammars like Head Driven Phrase Structure Grammar or similar notions in Type Logical Grammar). A witness for the type (6a) is given in (6b) – it includes a phonetic token – distinguished here from its associated phonological type in terms of spelling, contextual parameters – a situation sit0, a time time0, a speaker spkr0, addressee addr0, utterance time time1, an individual named Jo j0, and situations grounding the truth of the addressing, precedence, and naming conditions c10,c20,c30 – and the Austinian prepositional entity \[ \text{sit} = \text{sit0} \]
\[ \text{sit-type} = \text{Leave(j0, time0)} \].

c-params represents the type of entities need to instantiate a meaning:

Reconstructing the meaning/content relationship in terms of two fields c-params and content, originating in HPSG, allows in the current setting for the possibility of partially instantiating a content and maintaining this as the semantic representation of an utterance until a more detailed instantiation is available, an important possibility in conversation interaction, as we discuss in section 4.3.

Situation Semantics is one of the harbingers of dynamic semantics: the relational theory of meaning can be straightforwardly reconstructed as a specification of input/output contexts associated with uttering a given sentence. Indeed the paper (Barwise 1985) was one of the first to spell out a dynamic semantics for NPs, though (in common with most other works in the dynamic semantics tradition) does not spell out how to interface with
the discourse/utterance situation in the above sense. This ties in with a number of weaknesses which Barwise and Perry’s conception of context exhibits:

- No dynamics of indexicality is worked out (e.g. interaction between turn taking and structure of context) to deal with cases like (2a,b).
- It ignores metacommunication (the focus of sections 4.2. and 4.3.).
- In common with traditional speech act theory, it ignores conversational structure: to take two simple examples, the interpretation of the second ‘hi’ as a *counter greeting* derives from its position following an initial greeting. Similarly, the resolution of ‘No’ picks up in some way on the adjacent assertion:

(7) a. A: Hi. B: Hi  
b. A: I’m right, you’re wrong. B: No. I’m right, you’re wrong.

- Due to lack of calculus of constraints, it is not easy to use the relational theory of meaning as a logic which could allow an explicit account of which information can be derived from an utterance.

The utterance-based formulation of semantic theory pioneered by situation semantics was criticized as misguided by Kaplan (1989), Partee (1985); for a subsequent argument *contra*, along with a good review of Kaplan’s and related approaches, see Israel & Perry (1996). Indeed the utterance-based formulation has until recently had relatively little impact. Why? Putting aside sociological explanations, one might say that although the theory was intended for conversational language, the methodology and setting were that of the traditional isolated sentence, for which the pay offs do not seem sufficiently significant given the apparent theoretical investment. When these tools are applied to a dialogue setting, significant pay offs for this perspective emerge.

4. Meaning, utterances, and dialogue

4.1. Phenomena from spoken language

There has been growing interest in recent years in developing notions of context that can be used to semantically analyze linguistic phenomena characteristic of conversational language and to model dialogue interaction (see Ginzburg 1996b; Poesio & Traum 1997; Larsson 2002; Asher & Lascarides 2003; Ginzburg 2010). The *efficiency of language*, in the sense discussed above, and concomitant importance of context becomes yet more urgent an issue given how pervasive non-sentential utterances are in conversational settings – one word utterances are estimated to constitute between 30–40% of all utterances, 25% of these are propositional or interrogative, and hence involve significant contextual resolution (see e.g. Fernández 2006). In the remainder of this article I will focus on a number of semantic phenomena that occur in conversational interaction, whose analysis builds on the conceptual apparatus brought into prominence by situation semantics, in particular, the reification of utterances as real world events and the view of meanings as first class citizens of the ontology, not metatheoretical entities. As it turns out, this apparatus offers powerful tools that also offer solutions to old linguistic problems, viz. how to integrate into context non-semantic parallelism conditions, characteristic of ellipsis constructions, and grammatical gender agreement in anaphora.
The phenomena I consider here revolve around metacommunicative acts, which are rare in texts, but pervasive in dialogue. There are two main types of metacommunicative interactions – acknowledgements of understanding and clarification requests.

An addressee can acknowledge speaker’s utterance, either once the utterance is completed, as in (8a), or concurrently with the utterance as in (8b):

(8) a. Tommy: So Dalmeally I should safely say was my first schooling. Even though I was about eight and a half. Anon 1: Mm. Now your father was the the stocker at Tormore is that right? (British National Corpus (BNC), K7D)

b. A: Move the train ... 
   B: Aha
   A:... from Avon ...
   B: Right
   A:... to Danville. (Adapted from the Trains corpus)

Concomitantly with an utterance’s addressee acknowledging her understanding of an utterance, are a variety of facts about the utterance that potentially enter into the common ground. This is evinced, here for (9a), by the possibility of embedding them under a factive-presupposition predicate such as ‘interesting’. (9) exemplifies two classes of facts about the utterance that become presupposable, facts about the content of sub-utterances (9b–d) and also facts that concern solely the phonology and word order of the utterance (9e).

(9) a. A: Did Mark send you a love letter?
   b. B: No, though it’s interesting that you refer to Mark/my brother/our friend
   c. B: No, though it’s interesting that you bring up the sending of love letters
   d. B: No, though it’s interesting that you ask about Mark’s epistolary habits
   e. B: No, though it’s interesting that the final two words you just uttered start with ‘l’

A recurring theme since the Russell/Strawson dispute over definites has been the notion of presupposition failure (see article 41 (Heim) Definiteness and indefiniteness and article 91 (Beaver & Geurts) Presupposition). However, in interaction there is rarely failure as such. Rather, conversationalists’ mismatches lead to a clarification request (CR) – a query about an unclear aspect of a previous utterance – being posed. Natural Language allows for fine grained potential for CRs, using both sentential and non-sentential means. (10) illustrates a form-based taxonomy of CRs that covers virtually all of the CRs occurring in the BNC:

(10) a. A: Did Bo leave?
   b. WOT: B: Eh? / What? / Pardon?
   c. EXPLICIT: B: What did you say? / Did you say ‘Bo’ / What do you mean ‘leave’?
   d. LITERAL REPRISE: B: Did BO leave? / Did Bo LEAVE?
   e. WH-SUBSTITUTED REPRISE (SUB): B: Did WHO leave? / Did Bo WHAT?
   g. REPRISE FRAGMENTS: B: Bo? / Leave?
   h. GAP: B: Did Bo ... ?
   i. FILLER: A: Did Bo ... B: Win? (Table I from Purver, 2006)
In this taxonomy, four classes of contents were identified: they can be exemplified in the form of Explicit CRs:

(11) a. Repetition: What did you say? Did you say ‘Bo’?
   b. Clausal confirmation: Are you asking if Bo left? You’re asking if who left?
   c. Intended content: What do you mean ()? Who is ‘Bo’?
   d. Correction: Did you mean to say ‘Bro’?

In practice, though most CRs are not of the Explicit category. Many CR utterances are multiply ambiguous. The most extreme case are reprise fragments, which seems able to exhibit all four readings, though in practice 99% of cases found in the corpus study Purver, Ginzburg & Healey (2001) were either Clausal confirmation or Intended content. Ginzburg & Cooper (2004) and Ginzburg (2010) demonstrate that reprise fragments display parallelism on a syntactic and phonological level with its source. Clausal confirmation readings, on the one hand, and intended content and repetition readings, on the other, involve distinct parallelism conditions, suggesting that different linguistic mechanisms underlie the distinct understandings. Clausal Confirmation readings do not require phonological identity between target and source, as shown in (12a,b). Nonetheless, as (12c–f) show, they require partial syntactic parallelism: an XP used to clarify an antecedent sub-utterance $u_1$ must match $u_1$ categorically:

(12) a. A: Did Bo leave? B: My cousin? (Are you asking if BO, my cousin, left?)
   b. A: Did she annoy Bo? B: Sue? (Are you asking if SUE annoyed Bo?)
   d. A: Did he phone you? B: he? / #him?
   e. A: Did he adore the book. B: adore? / #adored?
   f. A: Were you cycling yesterday? B: Cycling?/biking?/#biked?

That repetition readings of RF involve (segmental) phonological identity with their source follows from their very nature (‘Did you say ...’). And this requirement also applies to intended content readings of RF:

(13) (i) A: Did Bo leave? B: Max? (cannot mean: intended content reading: WHO ARE YOU REFERRING TO? OR WHO DO YOU MEAN?)

The existence of syntactic and phonological parallelism in CRs across utterances is further evidence to that provided above in (9) that the notion of context we need is one that tracks non-semantic information associated with utterances, not merely content, presuppositions and the like. I will show that one way to capture this requirement is by defining contextual updates in terms of locutionary propositions, propositions constructed from utterances and the types that classify them. This idea has antecedents in the relational theory of meaning and in the Austinian conception of propositions, discussed in detail in article 35 (Ginzburg) Situation Semantics and NL ontology.

It should be emphasized just how central a phenomenon metacommunicative interaction is in interaction: a rough idea of the frequency of acknowledgements can be gleaned from the word counts for ‘yeah’ and ‘mmh’ in the demographic part of the BNC: ‘yeah’ occurs 58810 times (rank: 10; 10–15% of turns), whereas ‘mmh’ occurs 21907 times (rank:
36. Situation Semantics: From indexicality to metacommunicative interaction

Clarification Requests (CRs) constitute approximately 4–5% of all utterances (see e.g. Purver 2004; Rodriguez & Schlangen 2004). Moreover, there is suggestive evidence from artificial life simulation studies that the existence of CRs is not an incidental feature of interaction but a key component in the long-term viability of a language. Macura & Ginzburg (2006) and Macura (2007) show that when repair acts are a part of a linguistic interaction system, a stable language can be maintained over generations. Whereas, in a community endowed with a language that lacks CRification, as I refer to the interaction brought about by a CR, the emergent divergence among language users is so high that the language eventually dies out. Ignoring metacommunicative interaction, as has been the case for just about the entire tradition of formal semantics, means missing out one of the basic building blocks of linguistic interaction. Situation Semantics was itself complicit in this. However, the view of language it provides, with its reference to speech events as part of the semantic domain, and the reification of meanings provides important building blocks for a theory of metacommunicative interaction.

How then to integrate metacommunicative aspects into the semantic process? Such phenomena have been studied extensively by psycholinguists and conversational analysts in terms of notions such as grounding and feedback (in the sense of Clark 1996 and Allwood 1995, respectively) and of repair (in the sense of Schegloff 1987). The main claim that originates with Clark & Schaefer (1989) is that any dialogue move \( m_1 \) made by A must be grounded (viz acknowledged as understood) by the other conversational participant B before it enters the common ground; failing this CRification must ensue. While Clark and Schaefer’s assumption about grounding is somewhat too strong, as Allwood argues, it provides a starting point, indicating the need to interleave the potential for grounding/CRification incrementally; the size of the increments being an important empirical issue. From a semantic theory, we might expect the ability to generate concrete predictions about forms/meanings of metacommunicative interaction utterances in context. Such a characterization needs to cover both the range of possibilities associated with successful communication (grounding), as well as with imperfect communication – indeed it has been argued that miscommunication is the more general case (see e.g. Healey 2008). Thus, we can suggest that the adequacy of semantic theory involves the ability to characterize for any utterance type the update that emerges in the aftermath of successful grounding and the full range of possible CRs otherwise. This is, arguably, the early 21st century analogue of truth conditions. The update component of this criterion builds on earlier adequacy criteria that emerged from dynamic semantics’ frameworks (see article 38 (Dekker) Dynamic semantics). Nonetheless, these frameworks have abstracted away from metacommunication.

I now consider two general approaches that strive to develop semantic theories capable of delivering grounding conditions/CRification potential. The first approach, an extension of Discourse Representation Theory (DRT) (see article 37 (Kamp & Reyle) Discourse Representation Theory), aims at explicating inter alia the potential for acknowledgements and utterance-oriented presuppositions; the second approach, constructed from the start as a theory of dialogue per se, shows how to characterize CRification potential.

A crucial assumption both approaches bear in common, one that distinguishes them from other dynamic semantic work (e.g. Roberts 1996; Groenendijk 1998; Dekker 2004; Asher & Lascarides 2003), but one that seems inescapable if metacommunicative
interaction is to be tackled, is the need for *semantic distributivity*: given the fact that a single (public) input can lead to distinct outputs for each conversationalist, the effect of semantic operations can no longer be defined on a common ground *simpliciter*, but this needs in one way or another to be relativized across the conversational participants. This is exemplified in Turn Taking Puzzles (Ginzburg 1997) such as (14,15), where depending on who gets the turn, resolution possibilities for ellipsis vary:

(14) a. A: Who does Bo admire? B: Bo?
   b. Reading 1 (SHORT ANSWER): Does Bo admire Bo?
   c. Reading 2 (CLAUSAL CONFIRMATION): Are you asking who BO (of all people) admires?;
   d. Reading 3 (INTENDED CONTENT CLARIFICATION): (Who do you mean ‘Bo’?)

(15) a. A: Who does Bo admire? Bo?
   b. Reading 1: (SHORT ANSWER): Does Bo admire Bo?
   c. Reading 2: (SELF CORRECTION): Did I say ‘Bo’?

The relativization of context is what enables an account of the contrast between (14) and (15), sketched in section 4.3.: the semantic material necessary for ellipsis resolution in cases like (14c,d) can only emerge once a clarification request has been introduced by the addressee.

4.2. Acknowledgements, grounding, and micro conversational events

Massimo Poesio and David Traum and collaborators (e.g. Poesio & Traum 1997; Matheson, Poesio & Traum 2000; Poesio & Rieser 2009) have developed a framework known as PTT (not an acronym), which integrates a dynamic semantic framework (a version of DRT, Kamp & Reyle 1993) with a framework for representing conversational interaction inspired by speech act theory. One of the starting points of PTT is the assumption Speech Events as Semantic Entities (see (3b) above). On the basis of this, they assimilate the treatment of speech acts to the treatment of other events in DRT. Thus, conversational events can serve as the antecedents of anaphoric expressions, just like normal events. The standard DRT construction algorithm would assign to the text in (16a) an interpretation along the lines of (16b) (using the syntax from Poesio & Muskens 1997) for Discourse Representation Structures (DRSs) – a single DRS containing the merged propositional content of both assertions.). In contrast, Poesio and Traum hypothesize that upon hearing an assertion of that sentence, the common ground in a conversation would be roughly in (16c):

(16) a. A: There is an engine at Avon. B: It is hooked to a boxcar.
   b. [x,y,z,s,s': engine(x), Avon(w), s: at(x,w), boxcar(y), s':hooked-to(z,y), z is x ]
   c. [ce1,ce2]  ce1 : assert(A,B,[x,w,s| engine(x), Avon(w), s: at(x,w)]) ce2 : assert(B,A,[y,z,s'| boxcar(y), s':hooked-to(z,y), z is x])

(16c) records the occurrence of two conversational events, ce1 and ce2, both of type *assert*, whose propositional content are separate DRSs specifying the interpretation of
the two utterances in (16a). The discourse entities ce1 and ce2 can serve as antecedents both of implicit anaphoric references, e.g. in the case of ‘backward’ acts like answers to questions, and of explicit ones. Consider (17): this may be viewed as performing at least two functions here: implicitly accepting the option proposed in ce1, and performing a query. Indeed backward-looking acts – (for the backward/forward-looking dialogue act dichotomy see Core & Allen 1997) such as accept are all implicitly anaphoric to a previous conversational event (ce1 in this case), hence the assumption that conversational events introduce discourse markers just like normal events do.

(17) a. A: We should send an engine to Avon. B: Shall we use engine E3?
   b. \[ce1, ce2, ce3| ce1: open-option(A,B,[x,w,e\text{|engine(x),Avon(w),e\text{:send(A,B,x,w)}]}),
   ce2: accept(B,ce1) ce3: ask(B,A,[y,e\text{'}\text{|engine(y),E3(y),e\text{'}\text{:use(A,B,y)}]}])

In fact, as mentioned earlier, Poesio and Traum develop their theory on the basis of a strong and dynamicized version of Speech Events as Semantic Entities: an utterance is taken to be a sequence of micro-conversational events (MCEs). On this view, the discourse situation is updated not just when a complete sentence has been observed, but whenever a new event is observed. Psychological research suggests that such updates can take place every few milliseconds (Tanenhaus & Trueswell 1995), so that observing the utterance of a phoneme is sufficient to cause an update; but in practice PTT typically assumes that updates take place after every word. The incremental update hypothesis is not just motivated by psychological findings about incremental interpretation in sentential utterances, but by the fact that in dialogue many types of conversational acts are hardly, if ever, performed with full sentences. A class of non-sentential utterances that quite clearly lead to immediate updates of the discourse situation are those used to perform dialogue control acts such as take-turn, keep-turn and release-turn actions whose function is to synchronize the two participants in the conversation as to whom is holding the floor (Traum & Hinkelmann 1992) and acknowledgements. These conversational actions are sometimes performed by sentential utterances that also generate a core speech act (e.g., the second utterance in (17a)), but more commonly they are generated by single-word discourse markers like ‘mmh’, ‘okay’, ‘well’, ‘now’.

In PTT, lexicon and grammar are formulated as defeasible rules characterizing the update potential of locutionary acts. The motivation for defeasibility include psycholinguistic results about lexical access, e.g. work such as Onifer & Swinney (1981) demonstrating that conversationalists simultaneously access all meanings of ambiguous words. Lexical entries and syntactic rules link a precondition stated in terms of the phonological/syntactic characteristics of a micro-conversational event and a possible effect stated in terms of the possible meaning of that event. In particular, syntactic rules enable the construction of compound locutionary events, whose atomic constituents are the MCEs corresponding to utterances of individual words. Each locutionary act $la_i$ sets up the potential for a subsequent illocutionary act $il_j$ (one of whose) effects is to constitute an acknowledgement of $la_i$.

This provides the basis for a treatment of grounding and dialogue control particles. I illustrate this for ‘okay’ in its use as an acknowledgement particle; PTT assumes that locutionary acts generate – here in a causal sense introduced by Goldman (1970) – core speech acts. The lexical entry could be specified, roughly, as in (18), where $u$ represents a locutionary and $ce$ an illocutionary act respectively:
(18) lexical entry for ‘OK’: [u, ce] u: utter (A, ‘okay’), ce: acknowledge(A, ce), generate (u, ce) ]

(A highly simplified view of) the conversational score resulting from such an acknowledgement to an (ongoing) utterance by A in (19a) would be roughly as in (19b). This gives a schematic illustration of the emergence of utterance-related presuppositions – there are four micro-conversational events – each characterized in terms of its phonological syntactic, and semantic characteristics respectively – the events mce1, mce2 of uttering ‘an’ and ‘engine’ respectively, the compound event mce3 of uttering ‘an engine’ and the event mce4 of uttering ‘OK’; mce4 generates a core speech act, the acknowledgement of mce4;

(19) a. ... A: an engine B: OK ...
   b. [mce1, mce2, mce3, mce4, ce4]
      mce1: utter(A, “an”), cat(mce1) = det, mce1 \rightarrow \lambda P, Q[x]; P(x); Q(x)
      mce2: utter(A, “engine”), cat(mce2) = N, mce2 \rightarrow \lambda x engine(x)
      mce1 \prec mce2, Dtrs({mce1, mce2, mce3}), cat(mce3) = NP
      mce3 \rightarrow \lambda Q[x]; engine(x); Q(x), generate(mce3, ce3),
      mce4: utter(B, ‘okay’), cat(mce4) = intj,
      ce4: acknowledge(B, ce3), generate(mce4, ce4)]

4.3. CRification and meaning reification

The ability to both process and generate clarification questions is vital in all areas of Human-Computer Interaction, ranging from web search to expert systems. This is one reason why interest in integrating CRification into the semantic process is an issue that has attracted significant interest in computational semantic work (see Schlangen 2004; Purver 2006, DeVault et al. 2005). Above and beyond this, developing a theory which can predict the clarification potential of utterances, the possible forms and contents available for their clarification, is an important theoretical challenge. It represents one of the fundamental aspects of interactivity. To date, the main attempts in this direction have been made within the KoS framework (not an acronym) (Ginzburg & Cooper 2004; Purver 2004; Purver 2006; Ginzburg 2010), where a detailed treatment of the phenomena discussed in this section can be found. KoS is formalized in Type Theory with Records. What is crucial for current purposes about this formalism, which takes situation semantics as one of its inspirations, is that it provides access to both types and tokens at the object level. Concretely, this enables simultaneous reference to both utterances and utterance types, a key desideratum for modelling metacommunicative interaction. This distinguishes Type Theory with Records from Discourse Representation Theory, for instance, where the witnesses are at a model theoretic level, distinct from the level of discourse representations.

On the view developed in KoS, there is actually no single context, for reasons explained previously – instead of a single context, analysis is formulated at a level of information states, one per conversational participant. The type of such information states is given in (20a). I leave the structure of the private part unaanalyzed here, for details on this, see Larsson (2002). The dialogue gameboard represents information that arises from publicized interactions. Its structure is given in the type specified in (20b):
(20) a. TotalInformationState (TIS):
- dialoguegameboard : DGB
- private : Private

b. DGB =
- spkr : Ind
- addr : Ind
- c-utt : addressing(spkr,addr)
- Facts : set(Proposition)
- Pending : list(locutionary Proposition)
- Moves : list(locutionary Proposition)
- QUD : poset(Question)

In this view of context:

- The spkr/hearer roles serve to keep track of turn ownership.
- FACTS represents the shared knowledge conversationalists utilize during a conversation. More operationally, this amounts to information that a conversationalist can use embedded under presuppositional operators.
- PENDING: represents information about utterances that are as yet un-grounded. Each element of Pending is a locutionary proposition, a proposition individuated by an utterance event and a grammatical type that classifies that event. The motivation for this crucial modelling decision, which concerns the input to grounding and CRification processes and which carries on to the Moves repository, is discussed below.
- MOVES: represents information about utterances that have been grounded. The main motivation is to segregate from the entire repository of presuppositions information on the basis of which coherent reactions to the latest conversational move can be computed. For various purposes (e.g. characterizing the preparatory conditions of moves such as greeting and parting) it is actually important to keep track of the entire repository of moves.
- QUD: (mnemonic for Questions Under Discussion) – questions that constitute a “live issue”. That is, questions that have been introduced for discussion at a given point in the conversation and not yet been downdated. The role of questions in structuring context has been recognized in a variety of works, including Hamblin (1970), Carlson (1983), van Kuppevelt (1995), Ginzburg (1994), Ginzburg (1996a), Roberts (1996), Larsson (2002). There are additional ways for questions to get added into QUD, the most prominent of which is during metacommunicative interaction, as we will see shortly. Being maximal in QUD (max-qud) corresponds to being the current ‘discourse topic’ and is a key component in the theory.

The Dialogue GameBoard, then, constitutes the publicized context in KoS – taking into account that conversationalists’ DGBs need not be identical throughout. Work in KoS (e.g. Fernández & Ginzburg 2002; Fernández 2006; Ginzburg 2010) has shown that virtually all types of non-sentential utterance, ranging from short answers, propositional
lexemes (e.g. ‘yes’, ‘no’), through reprise fragments, can be analyzed as \textit{indexical} expressions relative to the DGB.

Context change is specified in terms of \textit{conversational rules}, rules that specify the \textit{effects} applicable to a DGB that satisfies certain \textit{preconditions}. This allows both illocutionary effects to be modelled (preconditions for and effects of greeting, querying, assertion, parting etc), interleaved with \textit{locutionary effects}, our focus here. In the immediate aftermath of the speech event \(u\), \texttt{PENDING} gets updated with a record of the form \[
\begin{bmatrix}
\text{sit} = u \\
\text{sit-type} = T_u
\end{bmatrix}
\] (of type \texttt{LocProp} (\textit{locutionary proposition})). Here \(T_u\) is a grammatical type that emerges during the process of parsing \(u\), as already exemplified above in (6). The relationship between \(u\) and \(T_u\) – describable in terms of the Austinian proposition (see (6) and article 35 (Ginzburg) \textit{Situation Semantics and NL ontology}) \(p_u = \begin{bmatrix}
\text{sit} = u \\
\text{sit-type} = T_u
\end{bmatrix}
\) can be utilized in providing an analysis of grounding/CRification conditions:

(21) a. Grounding: \(p_u\) is true: the utterance type fully classifies the utterance token.
b. CRification: \(T_u\) is weak (e.g. incomplete word recognition); \(u\) is incompletely specified (e.g. incomplete contextual resolution).

Thus, pending utterances are the locus off of which to read grounding/CR conditions.

Without saying much more, we can formulate a lexical entry for CR particles like ‘eh?’ (Purver 2004). Given a context that supplies speaker, addressee and a pending utterance the content expressed is a question querying the intended content of the utterance:

\[
\begin{bmatrix}
\text{PHON}: \langle \text{eh} \rangle \\
\text{CAT} = \text{interjection: syncat} \\
\text{spkr} : \text{IND} \\
\text{addr} : \text{IND} \\
\text{pending} : \text{utt} \\
\text{c2} : \text{address}(\text{addr}.\text{spkr}, \text{pending}) \\
\text{CONT} = \text{Ask}(\text{c-params}.\text{spkr}, \text{c-params}.\text{addr}, \lambda x \text{Mean}(\text{c-params}.\text{addr}, \text{c-params}.\text{pending}, x)) : \text{IllocProp}
\end{bmatrix}
\]

(22) is straightforward apart from one point – what is the type \textit{utt}. This actually is a fundamental semantic issue, one which, as we will see, responds to the \textit{underdetermination of information by interpretation} desideratum raised in section 1: what is the semantic type of \texttt{PENDING}? In other words, what information needs to be associated with \texttt{PENDING} to enable the formulation of grounding conditions/CR potential? The requisite information needs to be such that it enables the original speaker to interpret and recognize the coherence of the range of possible clarification queries that the original addressee might make.
Meanings – in the sense I discussed earlier of functions from contexts, which provide values for certain parameters (the contextual parameters), to contents – provide a useful notion for conceptualizing grounding/clarification potential (and were exploited for this purpose in Ginzburg 1996b). This is because the range of contextual parameters offers a possible characterization of the contextually variable and hence potentially problematic constituents of utterance content. Note though that if we conceive of meanings as entities which characterize potential sources of misunderstanding, the contextual parameters will need to include all open class sub-utterances of a given utterance type (i.e. including verb, common noun, and adjective, sub-utterances). This is a far cry from the 4 place indices of Montague and Kaplan, from the meanings envisaged by Barwise and Perry, and even from dynamicized meanings in dynamic semantics. (For experimental evidence about which lexical categories are viewed to be clarifiable see Purver 2004.)

Ginzburg & Cooper (2004) argue that, nonetheless, even radically context dependent meanings of this kind are not quite sufficient to characterize CR potential. One problem is the familiar one of grain. In terms of the concept or property that they represent, one would be hard pressed to distinguish the meanings of words such as attorney and lawyer. And yet, since knowledge of language is not uniform, it is clear that the clarification potential of the sentences in (23) is not identical. Which word was used initially makes a difference as to how the clarification can be formulated:

(23) a. Ariadne: Jo is a lawyer. Bora: A lawyer?/What do you mean a lawyer?/#What do you mean an advocate?/#What do you mean an attorney?

b. Ariadne: Jo is an advocate. Bora: #What do you mean a lawyer?/An advocate?/#What do you mean an advocate?/#What do you mean an attorney?

Other arguments derive from syntactic and phonological parallelism exhibited by non-sentential CRs (exemplified by (10f,g)) to their antecedent sub-utterance, and the existence of CRs whose function is to request repetition of (parts of) an utterance. Such CRs can, in principle, arise from any sub-utterance and are specified in terms of the utterance’s phonological type. Indeed the fact that any sub-utterance can, in principle, give rise to clarification motivates one one relatively minor enhancement to the standard grammatical representation. Instead of keeping track solely of immediate constituents, as is handled in formalisms such as HPSG the feature DTRS, we enhance the representation itself so it keeps track of all constituents. This is done by positing an additional, set valued field in the type definition of signs dubbed CONSTITUENTS, illustrated below in Fig. 36.3. In Ginzburg (2010), it is shown that this enhancement plays a key role in capturing cross-utterance parallelism, agreement, and scopal and anaphoric antecedency, though here I will only hint at the role it plays in formulating rules that regulate grounding and CRification.

The arguments provided hitherto point to the fact that PENDING must incorporate the utterance type associated by the grammar with the clarification target. This would have independent utility since it would be the basis for an account of the various utterance presuppositions whose source can only derive from the utterance type (see example (9)). In fact, we encounter here evidence for the assumption Speech Events as Semantic Entities: CRs typically involve utterance anaphoricity. In (24a,b) the issue is not what do you
mean by leaving or who is Bo in general, but what do you mean by leaving or who is Bo in this particular sub-utterance:

(24) a. A: Max is leaving. B: leaving?
   b. A: Did Bo leave? B: Who is Bo?

Taken together with the obvious need for pending to include values for the contextual parameters specified by the utterance type, Ginzburg (2010) argues that the type of pending combines tokens of the utterance, its parts, and of the constituents of the content with the utterance type associated with the utterance. An entity that fits this specification is the locutionary proposition defined by the utterance, as introduced before in (21).

With this in hand, I formulate in (25) a highly simplified utterance processing protocol, which interleaves illocutionary and metacommunicative interaction:

(25) **Utterance processing protocol**

For an agent A with DGB DGB0: if a locutionary proposition \( p_u \) = \[
\begin{bmatrix}
  \text{sit} = u \\
  \text{sit-type} = T_u
\end{bmatrix}
\]
is Maximal in PENDING:

(a) If \( p_u \) is true, update Moves with \( p_u \).
(b) Otherwise: introduce a clarification issue derivable from \( p_u \) as the maximal element of QUD; use this context to formulate a clarification request.

There are a small number of schemas that specify the possible clarification issues derivable from a given locutionary proposition \( p_u \). These include the issues ‘What did A mean by u1’ and ‘What did A utter in u1’, where A is the speaker provided in the contextual assignment represented in \( p_u \) and u1 is a sub-utterance of u. The hypothesis that the context has been incremented with such an issue is taken to be the explanation for how non-sentential CRs such as (10b,f,g) and (12) are interpretable domain independently.

To conclude, Fig. 36.3 offers a schematic illustration of how a single utterance – here of ‘Did Bo leave?’ – can lead to distinct updates among distinct participants at the ‘public level’ of context. In this case this arises due to differential ability to anchor the contextual parameters. The utterance u0 has three sub-utterances, u1, u2, u3, given in Fig. 36.3 with their approximate pronunciations. A can ground her own utterance since she knows the values of the contextual parameters, which I assume here for simplicity include the speaker and the referent of the sub-utterance ‘Bo’. This means that the locutionary proposition associated with u0 – the proposition whose situational value is a record that arises by unioning u0 with the witnesses for the contextual parameters and whose type is given in Fig. 36.3 – is true. This enables the ‘canonical’ illocutionary update to be performed: the issue ‘whether b left’ becomes the maximal element of QUD. In contrast, let assume that B lacks a witness for the referent of ‘Bo’. As a result, the locutionary proposition associated with u0 which B can construct is not true. Given this, B increments QUD with the issue ‘who was meant by A as the referent of subutterance u2’, and the locutionary proposition associated with u0 which B has constructed remains in Pending.
Token: u0

<table>
<thead>
<tr>
<th>di</th>
<th>bow</th>
<th>live</th>
</tr>
</thead>
<tbody>
<tr>
<td>u1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>u2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>u3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Speaker: Referent witnesses: wA = [spkr=A b=b0 ]

Addressee: Referent witnesses: wB = [spkr=A ]

phon: did bo leave
cat = S[+root]
constits: [u1 : aux u2 : NP u3 : VP]
c-params: [spkr: Ind b:Ind]
cont=Ask(spkr,?leave(b))

Grounding: Critification:

Moves:= <u0+wA,u0> Pending:= <u0+wB,u0>
MaxQUD:= ?leave(b) MaxQUD:= ?x.Intend(A,u2,x)

Fig.36.3: A single utterance gives rise to distinct Updates of the DGB for distinct participants.

5. Closing remarks

One of the innovative contributions of situation semantics has been the Relational Theory of Meaning, an utterance oriented approach to semantics, which naturalizes meanings as first class entities. The origins of this theory were somewhat philosophical, rooted in a desire for an ecologically realistic semantics, a semantics that takes seriously the resource bounded nature of situated agents. The tools that emerged in the wake of this stance have emerged in recent years as technically significant in the development of semantic analyses of actual conversational speech, specifically in the analysis of metacommunicative interaction, one of the constitutive features of conversation.

I would like to thank Robin Cooper and the Editors for many helpful comments on an earlier draft and Noor van Leusen for much helpful advice on finalizing this document.

6. References


37. Discourse Representation Theory

1. Introduction
2. DRT at work
3. Presupposition and binding
4. Binding in DRT
5. Lexicon and inference
6. Extensions
7. Direct reference and anchors
8. Coverage, extensions of the framework, implementations
9. References

Abstract

Discourse Representation Theory (DRT) originated from the desire to account for aspects of linguistic meaning that have to do with the connections between sentences in a discourse or text (as opposed to the meanings that individual sentences have in isolation). The general framework it proposes is dynamic: the semantic contribution that a sentence makes to a discourse or text is analysed as its contribution to the semantic representation - Discourse Representation Structure or DRS - that has already been constructed for the sentences preceding it. Interpretation is thus described as a transformation process which turns DRSs into other (as a rule more informative) DRSs, and meaning is explicated in terms of the canons that govern the construction of DRSs. DRT’s emphasis on semantic representations distinguishes it from other dynamic frameworks (such as the Dynamic Predicate Logic and Dynamic Montague Grammar developed by Groenendijk and Stokhof, and numerous variants of those). DRT is - both in its conception and in the details of its implementation - a theory of semantic representation, or logical form.

The selection of topics for this survey reflects our view of what are the most important contributions of DRT to natural language semantics (as opposed to philosophy or artificial intelligence).

1. Introduction
1.1. Origins

The origins of Discourse Representation Theory (DRT) had to do with the semantic connection between adjacent sentences in discourse. Starting point was the analysis of tense,