The Dynamics of Discourse Situations
(Extended Abstract)

Massimo Poesio
University of Edinburgh
Centre for Cognitive Science / HCRC
2 Buccleuch Place
Edinburgh EH8 9LW, UK
e-mail: poesio@cogsci.ed.ac.uk

Reinhard Muskens
University of Tilburg
Department of Linguistics
2 Warandelaan
5000 LE Tilburg, The Netherlands
e-mail: rmuskens@kub.nl

Abstract

The effects of utterances such as cue phrases, keep-turn markers, and grounding signals cannot be characterized as changes to a shared record of the propositions under discussed: the simplest (and arguably most natural) way of characterizing the meaning of these utterances is in terms of a theory in which the conversational score is seen as a record of the discourse situation, or at least of the speech acts that have been performed. The problem then becomes to explain how discourse entities are accessible. We consider three hypotheses about the dynamics of a speech act-based theory of the conversational score, and argue that they could be implemented with relatively minor modifications to the technical tools already introduced in theories such as Compositional DRT.

1   Introduction

The shared 'conversational score' in a conversation does not consist only of information about the propositional content of utterances. The participants in a conversation also share information about whose turn it is to speak, how what is being said fits in within the structure of the rest of the conversation, and whether what has been said needs acknowledging (Clark, 1996). Thus, the typical conversation consists not only of utterances performed to assert or query a proposition, but also of utterances whose role is to acquire, keep, or release a turn, to signal how the current utterance relates to what has been said before, or to acknowledge what has just been uttered (Poesio and Traum, 1997; Ginzburg, 1997). The linguistic tools used for these purposes include cue phrases such as *so* or (one sense of) *okay*; keep-turn signals such as *umm* or *wait*; and grounding signals such as *okay* again, *right* or *huhu*. Bunt (1995) proposed for these utterances the term DIALOGUE CONTROL ACTS.

Specifying the meaning of these expressions is a fundamental problem in the semantics of dialogues. In keeping with the assumptions of theories of the common ground such as DRT or DPL (Kamp and Reyle, 1993; Groenendijk and Stokhof, 1991), we are going to identify the meaning of an utterance with the way it modifies the conversational score; the problem we address in the paper is how to reconcile a model of the conversational score in which the update potential of these utterances can be explained with current views on how discourse entities become possible antecedents for anaphoric expressions.

2 The Basic Idea

A Simple Characterization of Discourse Situations

The effect of dialogue control acts is most naturally characterized in terms of a speech act-based theory. As a result, the problem of specifying the meaning of the expressions used to perform dialogue control acts was addressed in (Poesio and Traum, 1997) by proposing that the conversational score consists of a record of the speech acts performed during the conversation, i.e., a stripped-down characterization of what is called the DISCOURSE SITUATION in Situation Semantics (Barwise and Perry, 1983).[1] Furthermore, the approach adopted in (Poesio and Traum, 1997) was to continue using the tools introduced in DRT to specify the content of the conversational score, treating speech acts just like any other ordinary event. For example, whereas the ordinary DRT construction algorithm would assign to the text in (1) a DRS along the lines of (2), we hypothesize that the common ground in a conversation is more like (3), where we have adopted the syntax from (Muskens, 1995).

(1) A: There is an engine at Avon.
     B: *It* is hooked to a boxcar.

(2) $[x,w,y,u,s,s' |$ **engine**$(x),$**Avon**$(w), s :$ **at**$(x,w), $ **boxcar**$(y), s' :$ **hooked-to**$(u,y), u = x]$

(3) $[ce1,ce2|$

    $ce1 :$ **inform**$(A,B,[x,w,s|$ **engine**$(x),$ **Avon**$(w), s :$ ☐ **at**(x,w) ☐]),

    $ce2 :$ **inform**$(B,A,[y,u,s'|$ **boxcar**$(y), s' :$ ☐ **hooked-to**(y,u) ☐ , $u = x$])]

(3) records the occurrence of two CONVERSATIONAL EVENTS *ce1* and *ce2*, both of them **inform**s (we assume the speech act classification scheme from (Poesio and Traum, 1997)), whose propositional content are the DRSs specifying the interpretation of the two utterances in (1). As in (Kamp and Reyle, 1993; Muskens, 1995), a Davidsonian treatment of events is assumed, in which each event- or state-describing predicate **p** such as **hooked-to** or **inform** has an additional argument for the event (or state). We also follow the convention in (Kamp and Reyle, 1993) of writing $e : \mathbf{p}(x,y)$ rather than $\mathbf{p}(x,y,e)$ for these predicates.

The hypothesis that the shared information about the discourse situation also takes the form of a DRS has the advantage of simplicity, in that the same conceptual and technical tools can be used to characterize both the discourse situation and the content of utterances; but the real reason for adopting characterizations of the discourse situation along the line of (3) is that we can then use the accessibility mechanisms from DRT to describe references to aspects of the discourse situation. For example, treating speech acts in the same way that events are treated in DRT means that we can explain anaphoric reference to speech acts just as anaphoric reference to events is treated in DRT. Such references can be explicit or implicit, just as in the case of reference to events; both forms force us to assume that speech acts are explicitly recorded in the conversational score. An example of explicit reference is example (4a) (from (Webber, 1991)): given our hypothesis that the conversational score is like in (3), the reference to the first speech act in (4a) can be assigned the interpretation in (4b). We discuss examples of implicit references to speech acts and other cases of reference to the content of the discourse situation below.[2]

(4) a. A: The combination is 1-2-3-4.
         B: Could you repeat *that*? I didn't hear *it*.

    b. $[ce1,ce2,ce3|$
        $ce1 :$ **inform**$(A,B,[x|$**combination**$(x), x = $ *1-2-3-4*]),
        $ce2 :$ **request**$(B,A,[e,u|$ $e :$ **repeat**$(A,u), u = ce1])$,
        $ce3 :$ **inform**$(B,A,[|¬ [e',u'| e' :$ **hear**$(B,u'), u' = ce1]])]$

---

[1] More in general, the reader will notice below that the theory of meaning we adopt borrows many ideas from the work of Barwise and Perry as well as from (Cooper, 1992).

[2] Subordination relations between utterances such as those assumed in Grosz and Sidner's theory (Grosz and Sidner, 1986) or in SDRT (Asher, 1993) get interpreted as relations among speech acts–e.g., we can say that ce2 elaborates ce1. See below.

We adopt Muskens' reformulation of DRT in (Muskens, 1995), in which the expressions of DRT are interpreted in a typed logic which, besides types for eventualities and times, provides a type s for assignments ('states'); discourse entities are then treated as functions from assignments to values. Muskens introduces a relation $i[u_1, \ldots, u_n]j$ that holds between states i and j if j differs from i at most over the values assigned to discourse entities $u_1, \ldots, u_n$:

- $i[u_1, \ldots, u_n]j$ is short for $\forall v \ (u_1 \neq v \wedge \ldots \wedge u_n \neq v) \rightarrow (v(i) = v(j))$

- $i[]j$ is short for $\forall v v(i) = v(j)$.

DRSs are then interpreted as relations between assignments, and conditions as functions from assignments to truth values. The mapping from DRS expressions into type theory is provided by a translation function $\circ$ whose clause for DRSs, for example, is as follows:

$$[\mathsf{u}_1, \ldots, \mathsf{u}_n \mid \phi_1, \ldots, \phi_m]\circ \quad = \quad \lambda i \lambda j i [u_1, \ldots, u_n] j \wedge \phi_1 \circ(j), \ldots, \phi_m \circ(j)$$

(Muskens, 1995) does not make explicit provision for DRS occurring as arguments of extra-logical predicates, but the standard accessibility conditions of DRT can be added by requiring that the value of a discourse entity u be preserved within an embedded DRS K. This can be done by revising first of all the definition of the function $\circ$ translating DRS constructs into type theory to make it depend on the set of accessible discourse entities D:

$$[\mathsf{u}_1, \ldots, \mathsf{u}_n \mid \phi_1, \ldots, \phi_m]\circ_D = $$
$$\lambda i \lambda j i [u_1, \ldots, u_n] j \wedge \phi_1 \circ_{D \cup \{u_1, \ldots, u_n\}}(j), \ldots, \phi_m \circ_{D \cup \{u_1, \ldots, u_n\}}(j)$$

and then making the translation of DRSs embedded in conditions depend on the set of accessible discourse entities:

$$\mathbf{p}(\overline{x}, [\mathsf{u}_1, \ldots, \mathsf{u}_n \mid \phi_1, \ldots, \phi_m], \overline{y})\circ_D = $$
$$\lambda i \mathbf{p}(\overline{x}\circ_D, \lambda j \lambda k i \subseteq_D j \wedge j[u_1, \ldots, u_n]k \wedge \phi_1 \circ_{D \cup \{u_1, \ldots, u_n\}}(k), \ldots, \phi_m \circ_{D \cup \{u_1, \ldots, u_n\}}(k), \overline{y}\circ_D)$$

where we have used the notation $i \subseteq_D j$ to indicate that all the discourse entities in D assign the same value to i and j. This is sufficient to ensure that in (4b), the discourse entity *ce1* gets the appropriate value in the DRSs specifying the contents of speech acts *ce2* and *ce3*.

How Utterances Update the Discourse Situation

The advantage of adopting a compositional version of DRT is that we need not assume that (3) is constructed all at once; instead, we can hypothesize that it is built compositionally by concatenating the interpretations obtained separately for the two utterances, as in (5).

(5) $[ce1 \mid ce1 : \mathbf{inform}(A,B,[x,w,s \mid \mathbf{engine}(x), \mathbf{Avon}(w), s : \boxed{\mathbf{at}(x,w)}])];$

$[ce2 \mid ce2 : \mathbf{inform}(B,A,[y,u,s' \mid \mathbf{boxcar}(y), s' : \boxed{\mathbf{hooked\text{-}to}(y,u)}, u = x])]$

In turn, we can specify how these interpretations are constructed from the interpretations assigned to single utterances. We assume, as in (Poesio and Traum, 1997; Poesio, 1996) that the meaning of each word utterance specifies an update of the discourse situation, providing information about the occurrence of an utterance, about its syntactic classification, and its conventional meaning (i.e., its compositional meaning as specified in traditional formal semantics theories). An utterance of the noun *boxcar*, for example, results in the following update of the discourse situation, whereby the utterance of a new locutionary act *mce6* is recorded (we use the predicate **utter** to characterize locutionary acts), of syntactic category N and with semantic content $\lambda x \, [\mid \mathbf{boxcar}(x)]$.

(6) $boxcar \rightsquigarrow [mce6 \mid mce6 : \mathbf{utter}(A,"boxcar"), \mathbf{syn}(mce6) = n, \mathbf{sem}(mce6) = \lambda x \, [\mid \mathbf{boxcar}(x)]]$

Within this view of utterance interpretation, the update of the conversational score originated by utterances that generate dialogue control acts is easy to specify: discourse markers such as *okay* or turn-taking signals such as *umm* will result in updates like (7) and (8), respectively, where we have adopted again the speech act taxonomy from (Poesio and Traum, 1997) (note also that utterances like *okay* include implicit references to previous speech acts).

(7)   *okay* $\leadsto$ [*ce1*|*ce1* : **accept**(*A*,*ce2*)]

(8)   *umm* $\leadsto$ [*ce2*|*ce2* : **keep-turn**(*A*)]

The price to pay when adopting this view of utterance meaning is that semantic composition cannot be specified anymore in terms of operations that manipulate semantic objects; instead, we are inevitably lead to an inferential characterization of semantic composition along the lines of that proposed in (Pereira, 1990), where the combination of utterances in larger utterances and the specification of the meaning of these larger utterances are provided by inference rules. The particular approach adopted here involves defeasible inferences over the DRS obtained by concatenating the updates resulting from the utterances of single words (Poesio, 1996). For example, the second sentence in (1) results in the following incremental updates of the discourse situation, one for each word utterance:

(9)   [*mce1*|mce1:utter(a,"it"),cat(mce1)=pro, sem(mce1)=λp. [*u*|];p];
        [*mce2*|mce2:utter(a,"is"),cat(mce2)=aux, sem(mce2)=λP. P];
        .....
        [*mce5*|mce5:utter(a,"a"), cat(mce5)=det, sem(mce5)=λP. λp [*y*|];P(y);p];
        [*mce6*|mce6:utter(a,"boxcar"),cat(mce6)=n, sem(mce6)=λx [|boxcar(x)]]

Again as in (Poesio and Traum, 1997; Poesio, 1996), we assume that syntactic composition and semantic interpretation are the result of an inference process that results in hypotheses about how 'lexical' utterances combine together in phrasal ones, and these in larger phrasal hypotheses. The conventional meaning of these larger events is derived by combining the conventional meaning of the constituent events. For example, such inferences may result in hypothesizing that *mce5* and *mce6* are subconstituents of a larger event *mce7* of syntactic type NP, with the result that the following update is inferred (we use the symbol ↑ to indicate sub-constituency).[3]

(10)   [*mce7*| *mce7* : **utter**(*A*,"a boxcar"), **cat**(*mce7*) = *np*,**sem**(*mce7*) = λp [*y*|];[|boxcar(y)];p, *mce5* ↑ *mce7*, *mce6* ↑ *mce7*]

3 The Dynamics of Discourse Situations

The problem with (3) is that the dynamic properties of DRT apparently get lost: e.g., what makes the discourse entity *x* in the third argument of *ce1* in (3) accessible from within the DRS in the content of *ce2*? We consider three hypotheses about how the dynamics might work in discourse situations.

3.1 First Hypothesis: Implicit Dynamics

The first, most conservative hypothesis is that accessibility depends on discourse structure, i.e., that the propositional contents of speech acts that are part of the same 'discourse segment' are implicitly related. Assuming a function **sp** from a speech act to the next speech act in the same discourse segment (corresponding to Grosz and Sidner's 'Satisfaction-Precedes' function), we could hypothesize that discourse entities became accessible because the following holds.[4]

---

[3]Some readers might think that all of these complications are an excessive price to pay if all that we gain is the possibility of assigning a meaning to expressions that generate discourse control acts. There is a second advantage, however: the hypothesis that utterances result in updates of the discourse situation like those we have just presented also gives us a way to provide an account of the interpretation processes that take place when utterances are incomplete, as it is usually the case in spoken conversations (Poesio and Traum, 1997).

[4]This hypothesis about accessibility is similar to the one adopted in SDRT (Asher, 1993).

AX-SP $\forall ce1, ce2, A, B, C, D, K1, K2, i$
$[ce1 : \textbf{p}(A,B,K1)] \circ (i) \wedge [ce2 : \textbf{r}(C,D,K2)] \circ (i) \wedge \textbf{sp}(ce1) = ce2 \rightarrow (\forall j, k\, K2(j, k) \rightarrow \exists l\, K1(l, j))$
[for **p**, **r** in {inform,suggest, ...}, K1 and K2 DRSs]

3.2 Second Hypothesis: Resource Situations

The second hypothesis we consider derives from the treatment of definite descriptions in (Poesio, 1993). The key observation from that work is that in spoken conversations people often have to deal with more than one 'topic situation', i.e., states of the world: e.g., two people who are developing a plan are simultaneously discussing the world as it is now (say, the current decoration of the kitchen), the actions in the plan, and the world as it will result from the plan (the kitchen after redecorating). All of these situations (that we will call RESOURCE SITUATIONS, as in Situation Theory) are part of the common ground: that is, in order to interpret a referring expression, the listener must identify the resource situation with respect to which that expression should be interpreted. Some resource situations (e.g., the situation representing the world around the speakers) are completely specified at the beginning of the conversation; other situations (e.g., the one representing the state of the kitchen at a given point of the planning) are 'constructed' in the course of the conversation. The second hypothesis we consider is that all reference in dialogue is mediated by resource situations.

For simplicity, we assume here that resource situations are just states in the sense of Compositional DRT, and that their content is specified by DRSs. We use the notation s::K to indicate that resource situation s is characterized by DRS K, with the following interpretation:

(11) $s :: K = \exists i K(i, s)$

According to the second hypothesis, the conversational score includes not only a record of the occurrence of speech acts, but also a record of the current resource situations. Thus, in the kitchen example, the conversational score will be as follows:

(12)   [...*ce1, ce2, ... kitchen-sit, plan-sit1, ...*|

....
kitchen-sit **::** *[x1,x2,...|***window***(x1),* **table***(x2),....]*
plan-sit1 **::** *[y1,y2,...|***TV***(y1),,* **small-table***(y2),....]]*

We already noted that the resource situations may get updated by speech acts– for example, the situation resulting from the redecoration plan changes during the conversation. We are going to propose, more specifically, that these updates to the resource situations are specified by the content of the speech acts: for example, that in (1) we have an initial resource situation *rs1* for the world under discussion, and that the two speech acts extend this initial resource situation and the resource situation *rs2* resulting from this first update, respectively, as follows:

(13)   [*ce1,ce2,rs1,rs2,rs3*|
ce1 : **inform***(A,B, [x,w,s|* **engine***(x),* **Avon***(w), s :* **at***(x,w)](rs1,rs2)),*
ce2 : **inform***(B,A, [y,u,s'|* **boxcar***(y), s' :* **hooked-to***(u,y), u = x](rs2,rs3))]*

This hypothesis about the way the dynamics works (a variant of that proposed in (Poesio and Traum, 1997)) has the advantage that it gives us the tools to specify a theory of the process by which resource situations are chosen such as (Poesio, 1993); but in order for this implementation of the idea about resource situations to work, it is necessary to weaken Axiom 1 of Compositional DRT that specifies 'how many states' there must be in a model, so as to allow states to be the value of discourse entities while avoiding inconsistencies. The following version of the axiom only requires there to be an state j that differs from state i on the value of discourse entity (= object of type π) v for each object of type e.

AX1-WEAK $\forall i(state(i) \rightarrow \forall v_\pi, \forall x_e, \exists j(state(j) \wedge i[v]j \wedge v(j) = x))$

Anaphoric expressions such as pronouns and definite descriptions can now be taken to involve two presuppositional elements–the resource situation and the referent–rather than just one as in (Muskens, 1996):

(14)   *the* $\rightsquigarrow$ [u| u : **utter**(A,"*the*"), **cat**(u) = *det*,
             sem(u) = $\lambda P \lambda Q \lambda i \lambda j \exists k (P(i,k)(X(R)) \wedge Q(k,j)(X(R)))$]

where R is an unbound discourse entity denoting the resource situation that specifies the value for the discourse entity X denoting the referent of the definite description. Depending on the value assigned to R different antecedents become available: thus in (12) a reference to an object in the kitchen-situation will be interpreted by fixing R to *kitchen-sit*, whereas a reference to an object in the plan situation will be interpreted by fixing R to *plan-sit1*. References to objects in the discourse situation will be interpreted by letting the inner abstraction in (14) bind R, i.e., by setting $R = i$.

3.3   Third Hypothesis: Propositional Discourse Entities

Finally, we consider the hypothesis that discourse entities become accessible because the DRSs specifying the contents of speech acts themselves become the values of discourse entities. That this is the case is shown by examples such as:

(15)   A: There is an engine at Avon.
             B: Why is *that* important?

in which the antecedent of *that* appears to be 'the fact that the engine is at Avon', i.e., the content of the previous assertion by A (a DRS). But if we allow discourse entities denoting DRSs in the conversational score, the dynamics of the example (1) could be explained by assuming that the conversational score for (1) is as in (16):

(1)   A: There is an engine at Avon.
             B: *It* is hooked to a boxcar.

(16)   [*ce1,ce2,K,K',fspace0,fspace1,fspace2*| *fspace0* = [||],
             K = [x w s| **engine**(x), **Avon**(w), s : $\boxed{\textbf{at}(x,w)}$],*ce1* : **inform**(A,B,K), *fspace1* = *fspace0;K*,
             K' = [y u s'| **boxcar**(y), s' : $\boxed{\textbf{hooked-to}(y,u)}$, u = x],
             ce2 : **inform**(B,A,K'), *fspace2* = *fspace1;K'*]

According to this view of the dynamics, Grosz and Sidner's 'focus spaces' that, in the previous section, had been identified with situations, are identified with DRSs. As in the previously examined hypothesis, the assumption here is that at any moment there is a current 'focus space', initially empty (*fs0*). The content of each complete speech act gets added to the conversational score (thus becoming available for reference), and the focus space is also updated by concatenating the previous focus space to the new content. These updates are not unlike those proposed in recent work on modal subordination by Geurts (1994). As in that work, it is necessary to impose constraints on the assignments in order to allow discourse entities to have propositions as values.

4   (Preliminary) Conclusions

We are at too early a stage to draw definite conclusions as to which of these three hypotheses about the dynamics of discourse situations fit better the facts. For the moment, we will just remark that of the three hypotheses, the first is the simplest; the third, while requiring the most complex modifications to the semantics, relies on mechanisms that have to be introduced anyway to deal with propositional reference; and the second gives us an easy way to explain how certain interpretive mechanisms work.

References

Asher, N. 1993. *Reference to Abstract Objects in English*. D. Reidel, Dordrecht.

Barwise, J. and J. Perry. 1983. *Situations and Attitudes*. The MIT Press.

Bunt, H. C. 1995. Dialogue control functions and interaction design. In R.J. Beun, M. Baker, and M. Reiner, editors, *Dialogue in Instruction*. Springer Verlag, pages 197–214.

Clark, H. H. 1996. *Using Language*. Cambridge University Press, Cambridge.

Cooper, R. 1992. Three lectures on Situation Theoretic Grammar. Technical report, University of Edinburgh, Centre for Cognitive Science.

Geurts, B. 1994. *Presupposing*. Ph.D. thesis, University of Stuttgart.

Ginzburg, J. 1997. On some semantic consequences of turn-taking. In A. Benz and G. Jäger, editors, *Proc. of the Munich Workshop on Formal Semantics and Pragmatics of Dialogue*. University of Munich.

Groenendijk, J.A.G. and M.J.B. Stokhof. 1991. Dynamic Predicate Logic. *Linguistics and Philosophy*, 14:39–100.

Grosz, B. J. and C. L. Sidner. 1986. Attention, intention, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.

Kamp, H. and U. Reyle. 1993. *From Discourse to Logic*. D. Reidel, Dordrecht.

Muskens, R. 1995. Tense and the logic of change. In U. Egli, P. Pause, C. Schwarze, A. von Stechow, and G. Wienold, editors, *Lexical Knowledge in the Organization of Language*. John Benjamins, Amsterdam / Philadelphia, pages 147–183.

Muskens, R.A. 1996. Combining Montague Semantics and Discourse Representation. *Linguistics and Philosophy*, 19:143–186.

Pereira, F. C. N. 1990. Categorial semantics and scoping. *Computational Linguistics*, 16(1):1–10, March.

Poesio, M. 1993. A situation-theoretic formalization of definite description interpretation in plan elaboration dialogues. In P. Aczel, D. Israel, Y. Katagiri, and S. Peters, editors, *Situation Theory and its Applications, vol.3*. CSLI, Stanford, chapter 12, pages 339–374.

Poesio, M. 1996. Ambiguity and underspecification. In M. Bierwisch, R. Blutner, and R. van der Sandt, editors, *From underspecification to interpretation: Proceedings of the Berlin Workshop on Underspecification*. IBM Deutschland: Working Papers of the Institute for Logic and Linguistics, Heidelberg.

Poesio, M. and D. Traum. 1997. Conversational actions and discourse situations. *Computational Intelligence*, 13(3):309–347.

Webber, B. L. 1991. Structure and ostension in the interpretation of discourse deixis. *Language and Cognitive Processes*, 6(2):107–135.