A Multi-Agent Extension of DRT

Nissim Francez ∗ Jonathan Berg †
June 14, 1995

Abstract

This paper introduces MDRT, an extension of DRT to a multi-agent dialogue. The main issue handled is cross-speaker anaphora, whereby one speaker refers anaphorically to a discourse entity introduced by another speaker. We present a representation of multi-speaker discourses that allows a non-degenerate solution to the commitment problem: to what conditions about a discourse entity is a speaker committed when using cross-speaker anaphora?

1. Introduction

One of the interesting developments in the (formal) semantics of natural language is the emergence of theories for the representation of meanings of larger linguistic structures than a single sentence, the latter having been the target of classical semantics. Two notable examples are the (similar) theories presented in [5, 4]. Also related is the theory in [3], striving, in addition, towards compositionality. A focus of attention in all of these theories is the treatment of inter-sentential anaphora. In these theories, meaning is taken to be richer than truth conditions. The additional representational structure enables the explanation of the observation that logically equivalent sentences (i.e., having the same truth conditions) may display semantic differences when incorporated in a discourse.

A major aspect of all these theories is that of stressing the dynamicity of the semantics, a notable influence of the semantics of programming languages. We are interested here in enhancing this aspect, by considering multi-agent discourses, namely discourses with more than one participant. We refer to the resulting theory as MDRT (multi-agent DRT).

Not surprisingly, some inspiration might be drawn from a related phenomenon in the context of programming languages, namely distributed programming - programming collections of communicating computational agents. Here, agents are merely assertive, but their assertions (and denials) may be interpreted dynamically as changing representational schemes of discourse (referred to generically as Discourse Representation Structures (DRSs), following [5], the nomenclature of which we will use and extend). The reader is assumed to be familiar with [5], or with the presentation of DRT in [1]. A comprehensive treatment of DRT can be found in [7].

∗ Dept. of Computer Science, Technion (francez@cs.technion.ac.il)
† Dept. of Philosophy, University of Haifa (j.berg@uvm.haifa.ac.il)
As we show in this study, there are some problems that such an extension has to face, which were not present in DRT (in its current form, henceforth related to as uni-agent DRT). As a result, some possible extensions of the uni-agent DRT suggest themselves. Our results, following the theories mentioned above, focus on inter-agent anaphora, whereby one participant in a discourse refers anaphorically to discourse entities introduced by a different participant in (the same) discourse. A major question here is the following: What is the commitment of an agent when anaphorically referring to a discourse entity introduced by another agent?

We would like to stress that our interest here is in semantics. However, as multi-agent discourse is closer to daily conversation, it becomes much harder to separate semantic content (literal meaning) from the broader pragmatic content deriving from circumstances and conventions of use.

2. Preliminaries

We use ‘α’, ‘β’ etc., (possibly indexed) as meta-variables ranging over affirmative sentences of an extensional fragment of natural language (English in the examples considered). We do not present here an explicit definition of a specific fragment. Rather, we state some assumptions about the intended underlying fragment.

We use ‘A’, ‘B’, etc., as meta-constants denoting elements of a set S of assertive agents (referred to also as speakers, or discourse participants). The meta-variable S (possibly indexed) ranges over speaker names. Most of the presentation is in the framework of two speakers only, but generalizes straightforwardly to an arbitrary number of speakers.

A multi-agent discourse element has the form S : α, indicating that the speaker (which is the value of) S asserts (the value of) α. For example,

A: a man walks in the park.

is a discourse element, indicating that speaker A asserted ‘a man walks in the park’.

A multi-agent discourse is a (finite) sequence Si : αi, 1 ≤ i ≤ n, of discourse elements. Note that there are no restrictions on the distribution of speaker names in a discourse. Thus,

\[ A : α_1, B : α_2, A : α_3. \]
\[ A : α_1, A : α_2, B : α_3. \]
and
\[ A : α_1, A : α_2, A : α_3. \]

are all legal discourses. For readability, we abbreviate S : α, S : β to S : α β .

The (degenerate) case of a multi-agent discourse, in which S1 = S2 = \cdots = Sn is a uni-agent discourse, whereby it is clear that MDRT should subsume DRT.

To simplify the presentation, we assume a flat syntactic structure of a discourse, even though such a flat structure is not rich enough for applications such as discourse-level verb-phrase ellipsis [10, 8, 9].

We further make the following linguistic assumption, which at this point was not empirically tested, but allows focusing on the issues of interest in our study:

Assumption: For any sequence αi., 1 ≤ i ≤ n, of sentences (in the fragment) which is an acceptable discourse in uni-agent DRT, any arbitrary speaker-assignment S, 1 ≤ i ≤ n, yields a well-formed multi-agent discourse S : αi., 1 ≤ i ≤ n.

Note that this assumption involves no restriction on the distribution of personal pronouns (for the sake of preserving reference), For example:

A: a man walks in the park.
I love Mary. You love her too.

May yield a speaker distribution such as:

A: I love Mary. B: You love her too.

in which the reference of 'I' and 'you' changes correspondingly. Clearly, an adequate semantic theory of multi-agent discourse has to account for indexical reference to speakers, who are not considered discourse entities\(^1\). However, this is not only a matter of personal pronouns, since similar examples can be given with other indexical expressions, such as 'the person speaking/hearing'. This raises the general question of which features of a discourse are preserved by which transformations of speaker assignments.

Note also that the inverse transformation, by which a plain uni-agent discourse is obtained from a multi-agent discourse by omitting the speaker assignment, is more problematic. Consider the following examples, where '¬' is sentence-level negation.

**Example:** From A: α. B: ¬α., one would obtain α. ¬α., which under current DRT is a plain self-contradictory (uni-agent) discourse, not embeddable in any model.

There are possible interpretations of extensions of uni-agent DRT which accommodate such discourses in a non-degenerate way, reflecting a speaker's "change of mind," but we shall not deal with them here.

Consider another example:

**Example:**

A: I love a woman. B: I love a woman (too).

Omitting the speaker assignment yields the trivially redundant uni-agent discourse

\[ \text{I love a woman. I love a woman (too).} \]

which is naturally semantically equivalent to 'I love a woman' (ignoring pragmatic consideration such as a repetition for emphasis).

We would like to stress again that in this work we are interested only in (some) semantic features of discourse, completely ignoring pragmatic considerations such as coherence, cooperativeness among participants, and relevance (cf. Grice [2]). At this stage, the fragment considered is assumed not to include indexical reference to discourse agents; we defer treatment of that to another occasion.

### 3. The Basic Representation Problem

The main question to be answered by MDRT is this: what is a suitable multi-agent discourse representation structure (MDRS)? Several issues must be considered in resolving this question.

Let us start by reviewing the DRS\(^2\) attached to a simple uni-agent discourse:

\[ x \]
\[ \text{man}(x) \]
\[ \text{walk}(x) \]
\[ \text{whistle}(x) \]

A man walks in the park. He whistles.

In following the established practice in DRT, giving the semantics of this representation in terms of embedding, we have implicitly made two important choices:

1. Indefinite NPs are not treated as referring (i.e., what is known as their specific reading is not treated here). They can be incorporated into this framework in a rather simple way, by using the anchoring feature of DRT (see [6, 7]). This feature allows the direct connection of a discourse marker to an entity

\(^{1}\)That is, under their role of discourse participants. Of course, they may be mentioned explicitly in the discourse, thereby becoming discourse entities not under that role.

\(^{2}\)We use \[^\]p\[^\] to denote the predicate constant (in the formal DRT language) derived from the NL predicate \[^\]p\[^\]. Thus, \[^\]man\[^\] is derived from man.
in the model via some non-linguistic means. Every embedding has to respect this anchoring.

2. Universal generalization by means of indefinites, known also as their generic reading, is not treated either. Thus, we focus on the existential generalization reading of indefinites, following the mainstream of plain DRT.

Now suppose that a speaker assignment yields the following multi-agent discourse, which is obviously closely related to the previous one:

A: A man walks in the park.
B: He whistles.

Cross-speaker anaphora is a fundamental feature of multi-agent discourses, for which any successful choice of multi-agent DRS must account. The pronoun ‘he’ in the assertion of speaker B, in one of its interpretations\(^3\), refers anaphorically to the discourse entity corresponding to ‘a man’, introduced by speaker A.

Can we settle for the same representation for this discourse as for its uni-agent counterpart? Certainly such a representation accounts for cross-speaker anaphora.

We claim that such a solution is not adequate, for several reasons, to be given below. Basically, such a "centralized," cumulative representation would only fit discourses that are highly cooperative, so to speak, where the collection of agents can be viewed as one agent w.r.t to an external environment.

Instead, we shall opt for a representation by which to each speaker corresponds a distinct structure, a local DRS (LDRS), representing the contents of that speaker’s contribution to the overall discourse. An LDRS is subject to the same construction rules as a DRS and also contains the following new items:

- A label (from S) denoting the discourse participant with which the LDRS is associated.
- A new kind of discourse condition, \(\delta(m_1, m_2)\), where \(m_2\) is a discourse marker local to the LDRS, while \(m_1\) is a marker local to some other LDRS, labelled by a different participant label (a remote marker). This condition is referred to as a dependence relation, where \(m_2\) is said to depend on \(m_1\).

This representation is directly inspired by the notion of a local state in a distributed program, in contrast to that of a shared, global state, that would result from retaining the centralized uni-agent DRS representation also for the multi-agent DRT, the possibility rejected as described before.

For the example above our theory yields the following representation:

\[
\begin{array}{ccc}
A: & x & B: & y \\
& \text{man}(x) & \delta(x, y) & \text{whistle}(y) \\
& \text{walk}(x) & & \\
\end{array}
\]

The exact meaning of the dependency relation among discourse markers in different LDRSs, \(\delta(x, y)\), and its contribution to embeddability, are explained below. It replaces the more traditional ‘\(x = y\)’ coreference relation among markers in traditional DRT for the case of remote anaphora. Instead of making do with just one discourse marker, as in

\[
\begin{array}{ccc}
A: & x & B: & y \\
& \text{man}(x) & \delta(x, y) & \text{whistle}(y) \\
& \text{walk}(x) & & \\
\end{array}
\]

\(^3\) Of course, a deictic interpretation of the pronoun is available too, but, as already mentioned, is not to be treated here.
we introduce a second marker for much the same reasons motivating the use of conditions of the form `\( x = y \)` in uni-\textit{DRT}. See [11] for a discussion of this matter. The idea of using \textit{reference dependence} instead of \textit{reference identity} was raised also, in a different context, by [11].

We thus assume that some \textit{DRS construction rule} \( C \) is given for uni-agent \textit{DRT}, which in addition accommodates the introduction of dependencies in the case of remote anaphora. We further augment the rule with the following additional clause:

**Multi-\textit{DRS construction rule}:** If the next input sentence (in a given multi-agent discourse) is assigned speaker label \( A \), apply the rule \( C \) to that sentence w.r.t. the \textit{LDRS} associated with speaker \( A \).

### 4. The Speaker's Commitment Problem

Another important advantage of the proposed representation for \textit{MDRT} is that it allows for the attribution of different \textit{commitments} to the speakers participating in the discourse.

The basic question here is this: \textit{when} speaker \( B \) \textit{refers} anaphorically to a discourse entity via a marker \( x \) introduced by another speaker \( A \), to which properties of \( x \) (represented as conditions in \( A \)'s \textit{LDRS}) is speaker \( B \) committed?

Recall the previous example. Speaker \( A \) introduced an entity, represented in his \textit{LDRS} by marker `\( x \)` , with properties `\( \text{man}(x) \)` and `\( \text{walk}(x) \)`. Next, speaker \( B \) anaphorically uses that entity by means of a dependent pronoun \( y \), with `\( \delta(x, y) \)` , and additionally attributes to it the property `\( \text{whistle}(y) \)` . Is \( B \) committed to the entity being a man and walking in the park?

Note that \( B \) did not \textit{deny} these conditions, at least in the part of the discourse considered so far. So, the question is what to take as the default assumption.

A major factor to consider here is, that even though \( B \) did not deny so far any property attributed to the anaphorically referred-to entity by \( A \), he can do so in a natural continuation of the discourse. Consider a slightly different case:

\( A \): A man walks in the park. He whistles.
\( B \): He does not whistle.

The proposed \textit{MDRS} representation is the following:

<table>
<thead>
<tr>
<th>( A ):</th>
<th>( x )</th>
<th>( B ):</th>
<th>( y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{man}(x) )</td>
<td>( \delta(x, y) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{walk}(x) )</td>
<td>( \text{whistle}(y) )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This is intended to be interpreted as follows. Speaker \( B \) is \textit{committed to referring to the same discourse object referred to by speaker} \( A \) (a notion that still needs a proper definition), but is \textit{not committed to all} the conditions imposed on this discourse entity by speaker \( A \). In particular, speaker \( B \) need not be committed to additional conditions that speaker \( A \) might add in a \textit{later} contribution to the discourse, made \textit{after} \( B \)'s remote anaphoric reference. In fact, speaker \( B \) may even impose a condition \textit{contradicting} one imposed by speaker \( A \), as is the case in this example regarding \( \lambda z. \text{whistle}(z) \) as opposed to \( \lambda z. \neg \text{whistle}(z) \).

Note that if one adopted a centralized representation for multi-agent discourse, there would be no theoretical possibility of avoiding \textit{full commitment} (including "future conditions") by remote anaphora, as all conditions are shared. Clearly, such a degenerate solution would be inadequate. One advantage of the proposed interpretation is that the individual agent \textit{LDRS} construction remains
monotonic. Off cross-speaker anaphorical reference committed one to all previously imposed conditions on the referent, a speaker could not explicitly deny any of these conditions (as in the example), without contradicting himself. This does not seem to be the common pretheoretical interpretation of such cases.

For now, we make the following minimal assumption about whether a speaker is committed to conditions he does not explicitly deny: the speaker is commitment to (at least) the conditions directly arising from the introduction (into the LDRS) of the antecedent of the anaphoric pronouns. These conditions are referred to as the core conditions (on the appropriate discourse marker), and they are explicitly marked as such in the DRS by a box prefixed with the marker to which they belong. Our current assumption, not empirically tested, is that the core conditions do not include, for example, the modification effect of relative clauses or adjectives (if these are in the fragment treated). More research is needed for an exact determination of the core conditions, which may be affected by pragmatics. Similar classification under the term inaugural conditions appears in [12], where they are used for the characterization of "partial objects".

Thus, in the above example, speaker B is committed to \( \text{man}(y) \), which is assumed to be the core condition (see figure below), but not committed to \( \text{walk}(y) \). A future denial of the latter is still possible, preserving the speaker's consistency; but an explicit denial of the former will be considered inconsistent (on the part of the speaker).

Furthermore, we shall assume that in cases such as the following, where a speaker is explicitly committed to some conditions imposed by a partner, this commitment is represented by copying those conditions into the LDRS of the speaker.

A: A man walks in the park.
B: That is right, and he whistles.

This discourse generates the following MDRS:

\[
\begin{array}{|c|c|c|}
\hline
A : & x & B : & y \\
\hline
\text{man}(x) & \text{man}(y) \\
\text{walk}(x) & \text{walk}(y) \\
\text{whistle}(x) & \text{whistle}(y) \\
\hline
\end{array}
\]

In the example, the S-ellipsis in 'that is right' has been interpreted as a commitment (by speaker B) to all the conditions imposed so far by speaker A. There is another reading of this ellipsis, whereby it relates only to the most recent condition imposed by A. Note, however, that these are the only two possibilities. Speaker B cannot commit himself in a simple way, e.g., via ellipsis, to an arbitrary sub-collection of conditions imposed by A.

An interesting case is the following:

A: A man walks in the park.
B: (No, he does not.

Here we have a case of negative VPE (Verb-Phrase ellipsis), which requires copying the negation of the condition in A's LDRS to B's LDRS.

\[
\begin{array}{|c|c|c|}
\hline
A : & x & B : & y \\
\hline
\text{man}(x) & \text{man}(y) \\
\text{walk}(x) & \text{walk}(y) \\
\hline
\end{array}
\]

Note that the negation here is interpreted as VP-negation. It could be sentence negation as well. The interesting issue here is, that in such a situation, the indefinite NP
‘a man’ cannot be interpreted as an existential quantifier (i.e., non-specifically); it can only attain a specific meaning. Otherwise, the pronoun ‘he’ in B’s assertion would be interpreted incorrectly, as e.g., "no one walks in the park".

Finally, note that the following borderline MDRS turns out to be invalid, where $p$ is an arbitrary predicate.

\[
\begin{array}{c|c|c|c}
A: & x & B: & y \\
& p(x) & \delta(x, y) & \neg p(y)
\end{array}
\]

The problem here is that B denies all the properties of the marker introduced by A, so there is no sense in which he could speak (i.e., via anaphoric reference) about the same entity A. Here the minimal commitment is violated. This will always be the case when one LDRS contradicts all the conditions attributed to a marker depending on another marker (in another LDRS) of which all the denied conditions are asserted.

Finally, we would like to draw attention to two additional consequences of the proposed representation.

- The representation allows for cross-speaker Kataphora. This happens when a dependency is introduced where the dependent remote marker has not been introduced yet in the appropriate LDRS. This could occur naturally with definite NPs.

Example:

A: He is coming.

B: The Messiah is coming.

If such Kataphoric references are to be avoided, it should be on syntactic grounds.

- There is also a question as to which discourse markers are accessible as targets of remote anaphoric reference. The current representation views the local LDRS as a scope unit, hence only its top-level markers are accessible to other participants. This seems to agree with intuition. For example, it precludes cross-speaker variable binding by universal quantification, which causes box-splitting within an LDRS and generates inner scope units, the markers of which are not accessible, and the restrictor of which has to remain local to the same LDRS.

5. Embedding and speaker justification

Concern over the attribution of commitments strongly influences the desiderata of MDRT. In un-agent DRT, after obtaining a representation via a DRS one applies an elaborate notion of an embedding (into a model) in the definition of truth (in a model). The outcome for every DRS is binary: either it is true or it is not. For multi-agent DRT finer notions of embedding are needed, as the outcome concerns speaker’s justification, as seen above. In other words, the evaluation of a discourse can determine whether both speakers are right, one of them only, or even none of them. MDRT should support such outcomes.

First, we define an MDRS to be referentially well-formed if and only if the dependency relation $\delta$ is acyclic. We assume that we are dealing with referentially well-formed MDRSs.

A model for MDRT is the same as a DRT model. An embedding $e_i$ of an LDRS $L_i$ (of some given MDRS D) follows the same rules as a DRT-embedding, with the following additional clause for dependency conditions: If $L_i$ contains the condition $\delta(x, y)$ with $y$ belonging to some other LDRS $L_j$ (in D), then
the $y$-core conditions in $L_j$ have to be satisfied.

Note that there is a subtle issue here. As defined, an embedding of $L_i$ tests the conditions of remote markers on which local markers depend. However, it is not forced to choose the same value for the remote marker as would a local embedding of the remote $LDRS$. This is so in order to allow for remote anaphoric references even in case the whole discourse contribution of the other participant fails to be embeddable. When attempting a simultaneous embedding of all $LDRS$'s of $D$, an additional constraint on choice agreement is imposed (see below). It may be possible to impose this condition always and disallow anaphoric reference to antecedents that themselves fail to refer.

Now we are in a position to give a straightforward definition for the notions distinguishing MDRT from DRT: An $MDRS$ $D$ is $S$-justifying (for $S$ in $S$) if the $S$-labelled $LDRS$ is embeddable. $D$ is harmonious iff it is $S$-justifying for every $S$ in $S$ in such a way, that whenever a dependency condition $\delta(x, y)$ occurs in $LDRS L_i$ and $x$ is a marker in $L_i$, $e_i$ and $e_j$ agree on the choice for $y$. Other discursive notions, such as conflict (or argument), can now similarly be defined.

### 6. Conclusions

In this paper we made some first steps towards a multi-agent discourse representation theory. Our theory, MDRT, has been designed as an extension of DRT. The main issues handled and consequences drawn are as follows:

- *Cross-speaker anaphor* is studied as a central feature of MDRT.
- The issue of what is a speaker committed to when using cross-speaker anaphoric reference was identified as an important factor of the emerging theory. MDRT supports a non-degenerate solution, whereby a speaker commits only to some core conditions.
- A split $DRS$ ($MDRS$) is proposed as a representation supporting the required commitment.
- A finer notion of embedding, inducing *speaker justification* instead of global truth conditions, is proposed as the meaning of an $MDRS$.

### Acknowledgements

We would like to thank Shalom Lappin for useful discussions of our proposed extensions to DRT. The initiation of this work was partially supported by a grant from the joint fund from the Technion - IIT and Haifa university. It was completed by a partial support of the first author by a grant from the Israeli Ministry of Science and Technology "Programming languages induced computational semantics". The part of the first author was also partially supported by a grant from the Technion vice-president for research (Formal semantics for natural language) and by the fund for the promotion of research in the Technion.

### References


