# Tracing consensus formation in meetings: Annotation and incremental decision modelling in the MEET Corpus

**Ghazaleh Esfandiari-Baiat, Jens Edlund**
KTH Royal Institute of Technology
Stockholm, Sweden
geb@kth.se, edlund@speech.kth.se

**Abstract**

We present an incremental annotation scheme and discourse model designed specifically for the study of consensus formation in collaborative meetings. By grounding the representation in observable contributions and enforcing a strict no-lookahead principle, the model provides a tractable way to analyse how decisions emerge over the course of interaction. The resulting structures are intentionally minimal yet expressive enough to capture the evolving task state and support dynamic visualisation and replay of the decision process. A web-based reference implementation of the model demonstrates how the evolving decision state can be inspected and replayed during analysis. Together with a suitable corpus, this framework provides a practical foundation for investigating the multimodal dynamics of collaborative decision-making in professional meetings.

**Keywords:** speech corpus, collaborative meetings, ranking tasks, annotation, dialogue modelling

## 1. Introduction

Meetings are a central mechanism for collaborative decision-making in professional settings. Meeting participants often work toward a shared outcome through extended discussion, and it is important to examine not only the final outcome but also the interactional process through which consensus emerges over the course of the meeting.

To investigate this process in a controlled setting, we analyse a corpus of small-group meetings in which participants solve a collaborative ranking task. In each meeting, the group must agree on an ordered list of items with respect to a given criterion. This task defines a clear decision space in which the emerging consensus can be observed as participants progressively construct and adjust the ranking.

Given this structured decision space, our aim is to track how the group's decision state evolves throughout the meeting. At any point in the interaction, we maintain a current estimate of the emerging consensus based solely on the discussion observed so far. As participants introduce, question, and refine propositions, this estimate is updated to reflect the state of the discussion. Ideally, the estimate converges toward the consensus ultimately reached by the participants.

To support this analysis, we introduce an annotation scheme and a discourse model designed to represent the evolution of the decision state as the meeting unfolds. The framework interprets the interaction incrementally, allowing the discussion to be represented as a sequence of updates to the emerging consensus. In this way, the interaction can be analysed not only as a collection of utterances, but as a structured process in which propositions about the ranking are proposed, examined, and progressively stabilised.

The framework is guided by a small set of practical design principles that constrain both the annotation procedure and the interpretation of the interaction. In particular, annotations are assigned without lookahead, reflecting only the information available at the current point in the meeting. The scheme also deliberately captures only those aspects of the interaction that are essential for modelling the evolving decision state. Annotations are currently produced by human annotators, forming a human-in-the-loop process that supports inspection and correction of the evolving interpretation. The approach is intended to remain compatible with existing dialogue annotation traditions, including the ISO framework for dialogue act annotation, while focusing specifically on the task-oriented modelling of consensus formation.

The annotation scheme and discourse model are intended not only to represent the interaction, but also to support the investigation of the modelling process itself. During replay of a meeting, the evolving decision state can be inspected as the model updates its estimate of the emerging consensus. A web-based reference implementation demonstrates how the incremental state of the model can be visualised and replayed during analysis. This allows the effects of modelling assumptions to be examined while following the interaction over time, making the framework a practical tool for exploring how consensus formation can be captured through incremental discourse modelling.

The present paper makes three main contributions. First, we introduce an incremental annota-



*Proceedings of the 22nd Joint ACL - ISO Workshop on Interoperable Semantic Annotation and Representation (ISA-22)*, pages 62–67
12 May 2026. 

tion scheme designed specifically for collaborative ranking tasks in meetings. Second, we propose a minimal discourse model that derives the evolving decision state directly from these annotations while enforcing a strict no-lookahead interpretation. Third, we present a reference implementation that supports visualisation and replay of the decision process, enabling detailed investigation of how consensus emerges over the course of the interaction.

## 2. Background and related work

In this section we briefly point to prior work that is directly relevant to the present study, foregoing a comprehensive positioning within the many related research areas for reasons of space.

Research on meetings is closely related to the social psychology of small groups, which examined how groups coordinate activity and reach decisions through interaction (Davis et al., 1976). However, meetings constitute a more specific form of organised interaction than small groups in general. As noted by (Goffman, 1961), meetings are characterised by a sustained shared focus of activity among participants, which distinguishes them from more loosely structured forms of group interaction. Within the taxonomy of group tasks proposed by (McGrath, 1984), the meetings studied here fall into the category of decision-making tasks, in which participants must evaluate alternatives and converge on a joint choice.

Experimental studies of group decision making have often relied on structured ranking tasks, in which participants must agree on the relative importance of a fixed set of alternatives. A well-known family of such tasks are survival ranking exercises, where groups are asked to prioritise items that might help them survive in a hypothetical scenario. Participants must discuss the available items and collectively produce an ordered ranking based on their perceived utility. The meetings in the present corpus follow this experimental tradition and employ three such survival ranking tasks (Hall and Watson, 1970; Hare, 1952; Lafferty and Pond, 1974).

Several approaches to dialogue and discourse modelling have explored how the evolving state of an interaction can be represented incrementally as the dialogue unfolds. In spoken dialogue systems, incremental interpretation has long been used to maintain partial semantic representations that are updated as new input becomes available (Skantze and Edlund, 2004). Such approaches emphasise robustness and the ability to operate with no knowledge of future context and even without full knowledge of the past. At the same time, dialogue annotation frameworks have been developed to provide standardised ways of describing communicative acts in interaction, most notably the ISO framework for dialogue act annotation (Bunt et al., 2020). The present work aligns with these traditions while focusing specifically on modelling the emergence of consensus in a constrained decision-making task.

## 3. Annotation scheme

The annotation scheme design was guided by two methodological commitments. First, annotation is performed incrementally and without lookahead. Annotators record only the information that is available at the current point in the interaction, without using later context to reinterpret earlier contributions. The motivation is that participants themselves do not have access to future context, and our aim is to model the state of the decision process as it could plausibly appear to the participants at any given moment. The annotation procedure must operate under the same constraint.

Second, the scheme is deliberately selective. Only those aspects of the interaction that are directly relevant to the ranking task are annotated. The aim is not to reinvent a comprehensive representation of dialogue structure, but to produce a representation that remains closely aligned with the decision process that unfolds during the meeting.

The annotations are currently produced in the ELAN multimodal annotation environment. Separate tiers are used to represent different aspects of the interaction. TurnUnits provide the temporal segmentation of the speech signal, while additional tiers capture the task-oriented structure of the discussion. Of these, the proposition track, which references task entities (items and rank positions), is of particular relevance to this paper.

In the present scheme, any mention of a task entity (an ITEM, a RANK, or both) is annotated as a proposition. Each proposition represents the current association between ITEM and RANK introduced by the contribution, and a newly annotated proposition may modify or override the previously current proposition. ITEM and RANK values may each contain one or more values, reflecting cases where the speaker explicitly proposes multiple alternatives (e.g. "map should be first or second"). Annotators are trained to apply this principle consistently when identifying propositions in the dialogue. They record the surface form of the contribution according to the scheme, while resolution of references and inheritance of previously mentioned values are handled automatically during later processing. Interactional categories such as questions or acknowledgements are therefore not encoded directly; instead they can often be inferred from the form of the proposition.

Two special symbols distinguish underspecification from explicit unknowns. The underscore (_) indicates that the corresponding entity is not mentioned in the contribution. The question mark (?) indicates that the value is explicitly unknown or requested. Annotators are trained to apply this distinction when interpreting contributions: "_" reflects the absence of a mention, whereas "?" reflects an explicit request for, or removal of, a value. For example, prop(itemA, _) marks a contribution mentioning itemA without specifying a rank, whereas prop(itemA, ?) marks a contribution asking for the rank placement of itemA.

Where multiple alternatives are expressed, the corresponding argument may contain a set of values. For example, prop(map,[1,2]) represents a contribution proposing that the map may occupy either rank 1 or rank 2. Similarly, a contribution mentioning several items in relation to a rank may be represented as prop([itemA,itemB],3). Such sets are interpreted as explicit alternatives rather than as separate propositions.

Acknowledgements of propositions are encoded as the fully underspecified form prop(_, _), which represents an additional mention of the currently active proposition without introducing new task information. Negative acknowledgements function similarly to questions in that they remove the currently active association; they are therefore annotated using explicit unknowns (?). This treatment works well for the majority of cases observed in the data, but the precise handling of negation is still under investigation and the present representation may not capture all possible interactional patterns. Misunderstandings are not annotated explicitly. Each contribution is recorded according to its surface interpretation at the moment it occurs; discrepancies are resolved through later propositions in the dialogue.

## 4. Discourse model

The discourse model is designed as a minimal operational interpretation of the proposition annotation track. It treats the annotated propositions as observable constraints on the evolving task state of the meeting and derives the current decision state directly from these contributions as they occur. The model operates strictly incrementally and without lookahead: each new proposition updates the state of the model using only the information available at that point in the interaction. This mirrors the constraints imposed on the annotation process and ensures that the representation remains closely aligned with the unfolding discussion.

The model consists of three components. The proposition stack records the chronological sequence of annotated propositions and thus provides a complete incremental history of task-relevant contributions in the meeting. From this sequence, the model derives an Issue Under Discussion (IUD), which represents the currently active proposition after resolving underspecified references. Finally, a consensus ranking structure aggregates the successive IUD states into a running estimate of the emerging group ranking. Each component performs a minimal transformation of the previous one, preserving the incremental structure of the interaction while progressively deriving the task state.

Although the three components operate on the same propositions, they serve different roles in the model. The proposition stack functions as an event log of task-relevant contributions, the IUD represents the current conversational focus, and the consensus ranking tracks the evolving decision outcome.

### 4.1. The proposition stack

The proposition stack is the simplest component of the model. It is a chronological record of all propositions produced by the annotation procedure, stored in the order in which they occur in the interaction. Each entry corresponds directly to an annotated proposition and may optionally carry additional metadata such as speaker identity or temporal position in the dialogue.

The stack itself performs no interpretation: it simply records the observable task-related contributions as they appear. As a result, it functions as an event log of the decision process and provides the complete incremental history from which the remaining components of the model derive their state.

Because it preserves the complete incremental history of the discussion, the stack can also be replayed to reconstruct the evolving state of the model under different interpretations or analysis settings.

### 4.2. Issue Under Discussion tracker

The Issue Under Discussion (IUD) tracker represents the currently active proposition in the dialogue. While the proposition stack records the history of contributions, the IUD tracker maintains the present conversational focus derived from that history.

The term Issue Under Discussion is adopted from Larsson's dialogue management framework (Larsson, 2002), where it refers to the currently active conversational issue. In the present model, the IUD tracker operationalises this idea by maintaining a single proposition of the same form as the annotated propositions.

Each new entry in the proposition stack updates the IUD incrementally, resolving underspecified references by inheriting values from the previous state and allowing explicit unknown values to clear previously established associations.

The IUD tracker is updated whenever a new proposition is pushed onto the proposition stack. Resolution follows two simple rules.

Underspecified values, marked with “_”, do not overwrite the corresponding field in the current IUD and therefore inherit the previously active value. Explicit unknown values, marked with “?”, clear the corresponding field and remove any previously established association.

All other values overwrite the current value in the corresponding field. Through these operations the IUD tracker maintains a continuously updated representation of the item–rank association currently under discussion.

The apparent simplicity of this mechanism is deliberate: in task-oriented dialogue within constrained domains, such lightweight resolution has proven both robust and efficient compared to more complex approaches.

### 4.3. Consensus ranking

The consensus ranking represents the model’s current best estimate of the group’s decision state. While the proposition stack records the history of contributions and the IUD tracker represents the current conversational focus, the consensus structure aggregates these successive IUD states into a running representation of the emerging ranking.

The structure maintains two parallel representations of the ranking: one for unambiguous placements and one for ambiguous placements. The unambiguous representation contains rank–item associations in which a single item is assigned to a rank. The ambiguous representation captures cases in which the discussion leaves multiple alternatives open, either because an item may occupy several possible ranks or because several items are considered for the same rank. Together, these structures allow the model to represent both resolved and unresolved parts of the ranking at any given point in the meeting.

The consensus ranking is updated incrementally whenever the IUD tracker produces a new proposition. Updates follow a small set of consistency constraints. First, the model follows a last-mention-wins policy: a new placement overrides earlier placements that involve the same item or rank. Second, ranks are exclusive: assigning an item to a rank removes any previous assignments for that rank. Third, items are unique: when an item receives a new placement, any previous placements involving that item are removed.

Ambiguous propositions introduce alternative placements into the ambiguous representation. For example, a proposition such as prop(water,[1,2]) indicates that the item may occupy either of the two ranks, while prop([map,gun],2) indicates that either item may occupy the same rank. Such alternatives remain represented until subsequent propositions resolve the ambiguity or introduce new constraints.

Through these incremental updates, the consensus structure provides a compact view of how agreement gradually develops during the meeting, while still preserving the uncertainty that characterises intermediate stages of the decision process.

## 5. Examples and reference implementation

The discourse model described above has been implemented in a web-based reference system using web components. The current implementation fully supports the incremental processing of sequences of propositions and the equally incremental computation of the IUD and consensus ranking structures from these sequences.

The system also provides visualisation of these structures during analysis. Replay functionality is currently manual, but is under development and will support both step-wise replay based solely on the annotation sequence and time-aligned replay using the original audio recordings and time aligned annotations.

The following example illustrates how the model represents alternative proposals and their resolution during the discussion. In the dialogue, participants first establish tentative placements for two items. A subsequent contribution introduces an explicit alternative for the second rank by proposing two possible items for that position. The final contribution resolves this ambiguity by selecting one of the alternatives. Because the rank is not mentioned in the final utterance, the annotation uses the underspecified value ‘_‘, allowing the IUD to inherit the previously active rank. Table 1 shows the dialogue and the corresponding proposition annotations.

| | Utterance | Annotation |
|---|---|---|
| A | The map should be first. | `prop( 1,map)` |
| B | The knife should be second. | `prop( 2,knife)` |
| C | Knife or compass for second place. | `prop( 2, [knife, compass])` |
| A | Compass. | `prop(_, compass)` |

Table 1: Example dialogue fragment and corresponding proposition annotations.

The sequence of propositions produced by this exchange is processed incrementally by the discourse model. The resulting proposition stack, cur-

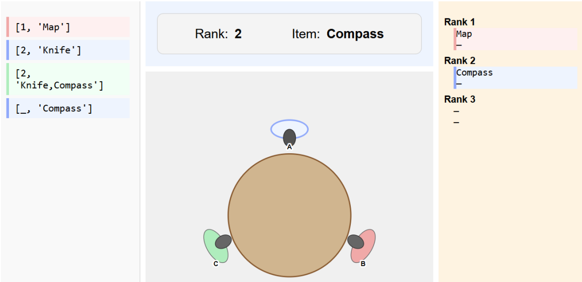


Figure 1: State of the discourse model after processing the dialogue shown in Table 1. The interface displays the proposition stack (left pane), the current IUD (centre top pane), and the evolving consensus ranking (right pane).

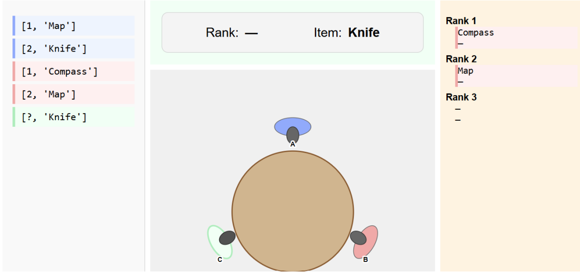


Figure 2: State of the discourse model after processing the dialogue shown in Table 2. The interface displays the proposition stack (left pane), the current IUD (centre top pane), and the evolving consensus ranking (right pane).

rent Issue Under Discussion (IUD), and consensus ranking are shown in Figure 1.

The following example illustrates how the model handles revisions of previously proposed rankings and the role of explicit unknown values. In this exchange, the participants first establish tentative placements for two items, after which a new proposal revises the ordering. The final contribution reopens the question of the placement of the knife. In this case the annotation uses the explicit unknown marker ‘¿, which clears the previously associated rank in the IUD. If the utterance were instead annotated using the underspecified value ‘_‘, the rank would be inherited from the current IUD state and the contribution would be interpreted as a renewed proposal for the previously active rank. The ‘¿ marker therefore allows the model to represent the contribution as reopening the placement decision rather than reinforcing or editing an earlier proposal. able 2 shows the dialogue and the corresponding proposition annotations.

| | Utterance | Annotation |
|---|---|---|
| A | The map should go first and the knife second. | `prop( 1, map)` `prop( 2, knife)` |
| B | No compass first, map second. | `prop( 1, compass)` `prop( 2, map)` |
| C | But what about the knife? | `prop( ?, knife)` |

Table 2: Example dialogue illustrating revision of previous proposals and the use of the explicit unknown marker ‘¿.

Processing the propositions derived from the dialogue in Table 2 yields the discourse model state in Figure 2, including the proposition stack, the current IUD, and the evolving consensus ranking.

## 6. Discussion

The modelling approach presented here is fully incremental and strictly constrained by the information available at each point in the interaction. This design reflects the methodological goal of studying consensus formation as an emergent process, rather than reconstructing it retrospectively using knowledge of the final outcome. By avoiding lookahead and global reinterpretation, the model approximates the informational conditions under which the participants themselves operate during the meeting.

The model is also narrow in scope. Rather than attempting to represent the full structure of dialogue, the annotation scheme and discourse model focus specifically on the ranking task that structures the meetings in the corpus. This restriction makes it possible to obtain a representation that is both robust and computationally tractable, while still capturing the aspects of the interaction that are most relevant to the collaborative decision process.

In this sense, the work draws on experience from other task-oriented dialogue domains, where lightweight incremental representations have proven effective for modelling constrained conversational tasks. Applying similar principles to collaborative human–human meetings provides a useful bridge between dialogue system research and the empirical study of group decision-making.

The resulting representation provides a structured account of the evolving decision process that can serve as a foundation for further analysis. Because the model tracks the state of the ranking incrementally over time, it can be aligned with other modalities present in the corpus, including acoustic-prosodic cues, gesture, and interactional dynamics across different meeting settings such as remote, co-located, or hybrid interaction.

Finally, the replayable structure of the proposition stack and discourse model enables dynamic visualisation of the decision process. When analysing recordings of real collaborative interaction, such visualisations provide valuable support for exploration and interpretation, allowing researchers to inspect how consensus develops over time. The reference implementation described above demonstrates how the incremental state of the model can be inspected and replayed, making it possible to evaluate modelling decisions and incorporate human expertise directly into the analytical process.

## 7. Conclusion and future work

This paper presented an annotation scheme and discourse model for studying consensus formation in collaborative meetings. The approach represents the evolving decision state of the meeting incrementally through a sequence of annotated propositions, an Issue Under Discussion tracker, and a consensus ranking structure. Together these components provide a minimal operational account of how group decisions emerge over the course of interaction.

Future work will focus on extending the reference implementation with full replay functionality and on applying the model to larger portions of the MEET corpus. The incremental representation also lays the ground for integrating additional modalities such as prosodic, gestural, and interactional signals in order to study how different communicative resources contribute to the formation of consensus in meetings.

## Acknowledgements

The results of this work will be made more widely accessible through the Swedish Research Council funded national infrastructure Språkbanken Tal (2023-00161_VR).

## 8. References

Harry Bunt, Volha Petukhova, Emer Gilmartin, Catherine Pelachaud, Alex Fang, Simon Keizer, and Laurent Prévot. 2020. The ISO standard for dialogue act annotation, second edition. In *Proc. of LREC'20*, pages 549–558, Marseille, France. European Language Resources Association.

J H Davis, P R Laughlin, and S S Komorita. 1976. The social psychology of small groups: Cooperative and mixed-motive interaction. *Annual Review of Psychology*, 27(1):501–541.

Erving Goffman. 1961. *Encounters: Two Studies in the Sociology of Interaction*. Encounters: Two Studies in the Sociology of Interaction. Bobbs-Merrill, Oxford, England.

Ernest James (Jay) Hall and W. H. Watson. 1970. The effects of a normative intervention on group decision-making performance. *Human Relations*, 23(4):299–317.

A. Paul Hare. 1952. A study of interaction and consensus in different sized groups. *American Sociological Review*, 17(3):261–267.

J. Clayton Lafferty and Alonzo William Pond. 1974. *The Desert Survival Situation: A Group Decision Making Experience for Examining and Increasing Individual and Team Effectiveness*. Human Synergistics, Plymoth, Michigan, US.

Staffan Larsson. 2002. *Issue-Based Dialogue Management*. Ph.D. thesis, Department of Linguistics, Göteborg University.

Joseph Edward McGrath. 1984. *Groups: Interaction and Performance*. Prentice-Hall.

Gabriel Skantze and Jens Edlund. 2004. Robust interpretation in the Higgins spoken dialogue system. In *Proc. of ROBUST 2004*, page 4, Norwich, UK. ISCA.