Non-Topical Coherence in Social Talk: A Call for Dialogue Model Enrichment

Alex Lưu
Brandeis University
alexluu@brandeis.edu

Sophia A. Malamud
Brandeis University
smalamud@brandeis.edu

Abstract

Current models of dialogue mainly focus on utterances within a topically coherent discourse segment, rather than new-topic utterances (NTUs), which begin a new topic not correlating with the content of prior discourse. As a result, these models may sufficiently account for discourse context of task-oriented but not social conversations. We conduct a pilot annotation study of NTUs as a first step towards a model capable of rationalizing conversational coherence in social talk. We start with the naturally occurring social dialogues in the Disco-SPICE corpus, annotated with discourse relations in the Penn Discourse Treebank (PDTB) and Cognitive approach to Coherence Relations (CCR) frameworks. We first annotate content-based coherence relations that are not available in Disco-SPICE, and then heuristically identify NTUs, which lack a coherence relation to prior discourse. Based on the interaction between NTUs and their discourse context, we construct a classification for NTUs that actually convey certain non-topical coherence in social talk. This classification introduces new sequence-based social intents that traditional taxonomies of speech acts do not capture. The new findings advocates the development of a Bayesian game-theoretic model for social talk.[1]

1 Introduction and Background

Social talk or casual conversation, one of the most popular instances of spontaneous discourse, is commonly defined as the speech event type in which "all participants have the same role: to be "equals;" no purposes are pre-established; and the range of possible topics is open-ended, although conventionally constrained" (Scha et al., 1986). Even though we do not establish any purposes in terms of information exchange or practical tasks, we do share certain social goal from the back of our mind when deciding to engage in a casual conversation. This work rests upon the assumption that casual conversations can be modeled as ***goal-directed rational interactions***, similar to task-oriented conversations, and therefore both of these types demonstrate Grice's Cooperative Principle, i.e. conversational moves are constrained by "a common purpose or set of purposes, or at least a mutually accepted direction" which "may be fixed from the start" or "evolve during the exchange", "may be fairly definite" or "so indefinite as to leave very considerable latitude to the participant" (Grice, 1975). A similar assumption is made in Grosz and Sidner (1986)'s discourse structure framework as it affirms the primary role of speakers' intentions in "explaining discourse structure, defining discourse coherence, and providing a coherent conceptualization of the term "discourse" itself." We adopt the following terminology from Grosz and Sidner (1986):

- utterances – basic discourse units.
- discourse segments – functional sequences of naturally aggregated utterances (not necessarily consecutive), each corresponding to a discourse segment purpose (DSP) – an extension of Gricean utterance-level intentions.

To account for conversational coherence, current models[2] of dialogue mainly focus on utterances within a topically coherent discourse segment, rather than **new-topic utterances** (NTUs), which begin a new topic not linguistically[3] correlating with the content of prior discourse. For example, the excerpt shown in Table 1 has two NTUs, utterances 119 and 123.

In terms of theoretical models, Asher and Las-

[1] The live version of this publication is located at https://osf.io/nvtkq/.

[2] Here we only consider the dialogue models that involve symbolic representation of discourse context (in comparison with, for example, end-to-end trained neural dialogue models).

[3] "Linguistically" means "via linguistic calculation at the meaning levels such as semantic or pragmatic."

Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, pages 118–133

Utt.	Simplified transcript
104-B	*And what 's the story with them*
105-B	*Are they still separated*
106-A	*Yes still separated*
107-A	*And Mummy was going she can't have children*
108-A	*Why Mummy it 's not her fault she can't have children*
109-A	*If he love her they could adopt*
110-A	*If he really wanted children of his own they [unclear speech]*
111-B	*I know*
112-B	*Sure he 's what forty odd five*
113-B	*Isn't he*
114-A	*Aye*
115-B	*Fucking hell*
116-B	*If he really wanted children he could 've had them long ago*
117-A	*That 's what I say*
118-B	*So uhm*
119-A	*Uh uh hold on*
120-A	*[unclear speech]*
121-A	*Think my mobile 's about to go*
122-A	*Ah it 's only John*
123-A	*Alright so how was your day*
124-B	*Not bad*

Table 1: An except, with indexed utterances, from dialogue P1A-095 in the SPICE-Ireland corpus (Kallen and Kirk, 2012) between two interlocutors A and B.

carides (2003)'s Segmented Discourse Representation Theory attributes conversational coherence to the existence of rhetorical relations between utterances, while Ginzburg (2012) and (Roberts, 1996/2012) propose that a conversational move is coherent if it is relevant to the Question Under Discussion. Computational models such as Belief-Desire-Intention (Allen, 1995, chapter 17) and Information State Update (Larsson and Traum, 2000) assume coherence to be a natural property of dialogues within a specific task domain. These models, both theoretical and computational, may adequately account for discourse dynamics of task-oriented conversations, where adjacent utterances tend to share a lot of linguistic material and speakers' intents are drawn from a narrow set of task-related goals. However, without any enrichment, they are not capable of handling the complexity of conversational coherence in social talk in which both speaker goals and utterances are less constrained. Specifically, all of these models treat NTUs as incoherent conversational moves.

This work, therefore, seeks to identify the constraints on new topics in casual conversations as a first step towards a model which is capable of rationalizing NTUs and accounting for conversational coherence in social talk. The main contributions of this paper are as follows. We introduce NTUs as a novel research object that is capable of advancing our understanding of the ***interactive*** and ***rational*** aspects of social talk. We propose an annotation strategy for exploring NTUs in naturally occurring dialogues. A pilot annotation study of NTUs in a significant amount of spoken conversation text led us to amend the available taxonomies of speech acts with new **sequence-based social intents** that shed light on non-topical coherence in social talk. These new findings feed into a framework for the ***Bayesian game-theoretic models*** that are capable of predicting the emergence of the newly identified intents and accounting for conversational coherence in social talk.

2 Methodology Overview

Before studying the interaction between NTUs and their discourse context, we need to locate them in instances of social talk. Riou (2015) handles a similar task by annotating every turn-constructional unit (TCU) in casual conversations with two topic-related variables:

- topic transition vs. topic continuity.
- stepwise vs. disjunctive transition (Jefferson, 1984) if the TCU is annotated as a transition.

The TCUs triggering disjunctive transitions are intentionally equivalent to NTUs and the corresponding transitions can also be called **disjunctive topic changes**[4] (DTCs), i.e. conversational moves whose linguistic representation is an NTU. To perform the annotation task in Riou (2015), the annotators completely rely on their own intuition rather than guidelines.[5] This negatively affects annotation reliability, especially for topic transition cases, which are much less frequent in the studied data.

[4]Sharing Jefferson's characterization of troubles-telling exit devices in that the new topic "does not emerge from [prior talk], is not topically coherent with it, but constitutes a break from it" (Jefferson, 1984), and comparable to TOPIC-SHIFT (Carlson and Marcu, 2001) in RST Discourse Treebank.

[5]This is because the author aims to investigate the linguistic design of topic transitions and therefore cannot give the annotators the linguistic description of these transitions. Otherwise, she would face the risk of circularity in her study.

To improve the reliability and rigor of NTU detection, we approach the task reversely: we first annotate content-based coherence relations between utterances and then identify NTUs as those utterances that bear no coherence relation to the content of prior discourse. This approach shares certain features with the integration of new utterances in free dialogues presented in Reichman (1978): if a new utterance is not covered by the current conversational topic, the hearer can expand the current topic to cover it, or connect its topic with the current topic using a semantic relation from a predefined set. This similarity reflects the following view of discourse coherence: "[a discourse is] coherent just in case (a) every proposition (and question and request) that's introduced in the discourse is rhetorically connected to another bit of information in the discourse, resulting in a 'single' connected structure for the whole discourse; and (b) all anaphoric expressions can be resolved"; and therefore, "[a] discourse is incoherent whenever there's a proposition introduced in the discourse which doesn't seem to be connected to any of the other bits of the discourse in any meaningful way." (Asher and Lascarides, 2003, p. 4).

The main difference between Reichman (1978)'s model of topic shift and our work is that the former allows the total shift relation, the succeeding topic of which is totally new, only when all of the preceding topics have been exhausted and closed, while we do not impose any constraints on the nature of DTCs. We assume that interlocutors are coherent in naturally occurring conversations (wherein incoherent moves need convincing evidence). Analyzing the coherence of a conversation, we put ourselves in conversational participants' shoes and rely on our communicative competence to identify all possible DSPs that account for the relevance of each conversational move. We are interested in the cases where an identified DSP cannot be assigned to a pre-existing coherence relation. We hypothesize that the pre-existing coherence relations account for topical coherence (i.e. talk-about), but not non-topical coherence such as interactional coherence (i.e. talk-that-does) (Clift, 2016, p.92).

3 Annotating Coherence Relations

We start with the casual telephone dialogues in the Disco-SPICE corpus[6] (Rehbein et al., 2016), based on the SPICE-Ireland corpus[7] (Kallen and Kirk, 2012), in which discourse relations – triples consisting of a discourse-level predicate and its two arguments – are annotated with the CCR (Sanders et al., 1992) and the early version of the PDTB 3.0 (Webber et al., 2016) schemes. We ignore the CCR annotations in favour of the PDTB 3.0-based annotation because the latter covers more discourse relations in the corpus, including:

- explicit discourse relations between any two discourse segments (whose predicate is an explicit discourse connective such as *"because"* or *"however"*).
- implicit/AltLex relations between utterances given by the same speaker (whose predicate is not represented by an explicit discourse connective but can be inferred or alternatively lexicalized by some non-connective expression, respectively).[8]
- entity-based coherence relations (EntRel) between adjacent utterances given by the same speaker (whose predicate is an abstract placeholder linking two arguments that mention the same entity).

In the excerpt shown in Table 1, utterances 104 and 105 are two arguments of an implicit relation that can be realized by a connective *"in particular"*, while 121 and 122 are the arguments of an entity-based relation that is signaled by the pronoun *"it"*.

We enrich Disco-SPICE with SPICE-Ireland's original pragmatic annotation, consisting of Searlean speech acts (Searle, 1976), prosody, and quotatives among others. This information is helpful in identifying, for example, the quote content, or speech act *query*, i.e. asking for information, even in declarative clauses.

We use the latest version of the PDTB 3.0 taxonomy of discourse relations (Webber et al., 2019), and annotate the instances which are not covered in the Disco-SPICE corpus, such as:

- implicit/AltLex discourse relations between utterances given by different speakers.
- entity-based coherence relations between adjacent utterances given by different speakers.
- entity-based coherence relations between non-adjacent utterances.

[6]This corpus is unique as it is publicly accessible, and highly relevant to our work in that the discourse relations are annotated in a significant amount of spoken conversation text.

[7]This corpus can be obtained upon request to its directors.

[8]Here we make an assumption that the same annotation strategy is applied to both implicit and AltLex discourse relations, since AltLex relations must first be identified as implicit ones (Webber et al., 2016).

Specifically, if a relation is not entity-based, it will be labeled with a sense in the PDTB 3.0 sense hierarchy. Annotators are encouraged to choose the most fine-grained labels. For example, *expansion.equivalence* is preferred over *expansion* for an *expansion.equivalence* relation, although both are acceptable. In total, there are 53 sense labels available for explicit/implicit/AltLex discourse relations.

We also enrich our repertory of content-based coherence relations with additional semantic relations from ISO 24617-8 and ISO 24617-2, which take care of the interactive nature of dialogue:

- functional dependence relations characterizing the semantic dependence between two dialogue acts due to their communicative functions (cf. adjacency pairs in Conversation Analysis)[9], named after the first pair part:
  - information-seeking: *propositionalQ*, *checkQ*, *setQ*, *choiceQ*.
  - directive: *request*, *instruct*, *suggest*.
  - commissive: *promise*, *offer*.
  - social obligation management: *apology*, *thanking*, *greeting*, *goodbye*.
- *feedback* dependence relations connecting a stretch of discourse and a response utterance that provides or elicits information about the success in processing that stretch.
- additional entity-based coherence relations relating to other communicative functions such as *topic closing* (as a discourse structuring function) and *completion* (as a partner communication management function).

In Table 1, utterances 105 and 106 are two arguments of a *propositionalQ* functional dependence relation, while 109 and 111 are the arguments of a *feedback* relation.

It is worth noting that the argument order of annotated coherence relations is chronological, i.e. the second argument always appears after the first argument in the conversational flow.

We aim at annotating coherence relations that cover as many utterances as possible (rather than exhaustively annotating every relation), adding notes to the ones that are not very clear and therefore can be considered non-existent in the next step – NTU identification. In case of multiple relations available to the same pair of arguments, annotating just one relation is sufficient. Table 2 shows the key statistics of the annotation in this work, performed solely by the student author (see further details of the annotation in Appendix A).

[9]Examples of adjacency pairs are *greeting - greeting*, *question - answer*, *request - grant/refuse*, etc.

10 dialogues - ***2,719*** utterances	
Inherited from Disco-SPICE: ***1,273*** coherence relations (***158*** entity-based)	
Newly annotated: ***1,870*** coherence relations	
implicit discourse relations	***10***
entity-based discourse relations	***1,490***
functional dependence relations	***324***
• information seeking	*291*
• directive	*4*
• commissive	*1*
• social obligation management	*28*
feedback dependence relations	***487***

Table 2: Statistics of coherence relation annotation.

As seen in Table 2, the ratio of the coherence relations inherited from Disco-SPICE to the newly annotated ones is $1,273/1,870 \approx 2/3$, which means that using Disco-SPICE saves us a considerable portion of annotation workload. While this efficiency is optimal for a pilot study, it does not provide the full picture of our proposed annotation task. We plan to use this study's annotation guidelines to conduct a full-blown annotation project on the data set[10] composed by Riou (2015), aiming at (1) performing in-depth empirical studies such as detailed analyses of the distribution of annotated relations and annotation disagreements, and (2) enriching the linguistic resources for studying dialogue coherence. In addition, the results of this study can serve as an assessment of the reliability of Riou (2015)'s annotation methodology.

4 Identifying NTU Candidates

Based on both inherited and newly annotated relations described in Section 3, excluding those relations noted as "not very clear", which account for less than *3%* of the newly annotated relations, we heuristically identified *72* candidates for NTUs, each of which is:

- not the first utterance of a dialogue,
- the first utterance token of the first argument of some coherence relation,

[10]This data set includes 15-min extracts of 8 conversations from the Santa Barbara Corpus of Spoken American English (Du Bois et al., 2000). The advantage of this data set over Disco-SPICE is that its audio files are publicly accessible, which is invaluable for our annotation.

- not part of 2nd argument of another relation,
- not in the dialogue span of another relation.

5 Identifying NTUs and Patterns of DTCs

An NTU candidate identified in Section 4 is valid only if there is no a content-based coherence relation with respect to prior discourse, which can be missed or annotated as "not very clear" in Section 3. To separate genuine NTUs from other NTU candidates, we carry out a more detailed inspection. Specifically, the following pieces of information are further annotated for each NTU candidate:

- the immediately preceding topic.
- the current topic, its focused entity[11], and its information status, i.e. given-new w.r.t. discourse/hearer (Prince, 1992; Birner, 2006).
- the interlocutors involved in content, if any, and their roles (speaker/hearer).
- the links between the current topic and:
  - the pre-dialogue common ground.
  - the utterance situation (time and space).
  - the content of prior discourse.

We were able to single out *38* true cases of NTUs, roughly *50%* of NTU candidates, which contain discourse-new topics and new focused entities. Based on the annotated information about the interaction between the NTUs and their discourse context, we identified the following patterns of DTCs (see detailed examples in Appendix B):

- Grosz and Sidner (1986)'s true interruption.
- forgotten topic (when the speaker cannot articulate the topic she intents to talk about).
- the first topic after greeting.
- goodbye-initialized topic (when saying goodbye opens a new discussion thread).
- interlocutor-decentric move (from a topic focusing on one of the interlocutors).
- interlocutor-centric move:
  - interlocutor-centric return (from a topic not focusing on the interlocutors).
  - interlocutor-centric switching (from a topic focusing on one interlocutor to a topic focusing on the other).
  - urgent interlocutor-centric topic in extra-linguistic utterance situation (when the speaker suddenly prioritizes an urgent topic related to one of the interlocutors).
  - speaker-centric distraction (an off-track topic focusing on the speaker).
  - speaker-centric wrap-up (when the attempt to wrap up the conversation opens a new discussion thread).
  - hearer-centric related topic (from a topic not focusing on interlocutors).
- cushioning topic (from interlocutor-decentric to interlocutor-centric) - topic immediately relevant to an interlocutor's life.

The presence of cushioning topics implies that the speaker may plan, at least, "two steps ahead", including:

- the interpretation the hearer may have, and
- the potential of topic extension based on that interpretation.

In addition, the patterns of goodbye-initialized topic and speaker-centric wrap-up can elicit better insight into the findings in Gilmartin et al. (2018) about the extended leave-taking sequences.

6 Classifying NTUs

The patterns of DTCs identified in Section 5 (except for Grosz and Sidner (1986)'s true interruption and the forgotten topic, covering *7* identified instances of NTUs) show that non-topical coherence, sustained or built by DTCs, is created via sequential adjustment of the distances between the active conversational topic and each interlocutor. This adjustment seems to be constrained by the relational work between the interlocutors, i.e. the social aspect of the conversations, rather than the content-based relevance.

Based on the interlocutors' intents, a simple version of the classification of NTUs in social dialogues, covering 31 identified instances of NTUs, can be proposed as below:

- socially initialized topic (the first topic after greeting) - *2* instances.
- topic merely motivated by changing social focus (urgent interlocutor-centric topic in extra-linguistic utterance situation, speaker-centric distraction) - *3* instances.
- topic merely motivated by changing the degree of relevance of social domains (interlocutor-decentric move, cushioning topic, interlocutor-centric return) - *9* instances.
- topic motivated by changing both social focus and the degree of relevance of social domains (generally embodied in the other patterns of

[11] Inspired by the ideas of focus of attention and local coherence in Grosz et al. (1995).

DTCs) - *17* instances.

This classification introduces new sequence-based social intents[12] that traditional taxonomies of speech acts do not capture as the social intents proposed in these taxonomies, if any, do not demonstrate the sequential dynamics of the relational work between the interlocutors (e.g. ISO 24617-2's social obligation management functions, Klüwer (2011)'s dialogue acts for social talk, or van der Zwaan et al. (2012)'s social support categories).

These newly found intents, characterizing non-topical coherence in social talk, convincingly demonstrate social talk as a sophisticated form of goal-directed rational interactions rather than a random walk through loosely connected topics. This shows real promise and new perspectives for research in dialogue modeling. We hypothesize that a workable dialogue model for social talk needs to explicitly handle all of the key aspects of goal-directed rational interactions.

7 Toward a Game-theoretic Model

To formally capture the interactive and rational aspects of social conditioned language use in conversation, recent work such as Iterated Best Response (Franke, 2009), Rational Speech Act (Frank and Goodman, 2012), and Social Meaning Game (Burnett, 2019) pairs Lewis (1969/2002)'s signaling games with the Bayesian approach to speaker/listener reasoning. In essence, these models formalize Gricean inference by predicting:

Speaker behavior: the probability $P_s(o|h, C_s)$ that the speaker uses the observed linguistic value o to convey hidden meaning h in the speaker's context model C_s is a function of $U_s(o, h, C_s))$, the utility of o in C_s given the speaker's desire to communicate h.

- $P_s(o|h, C_s) \propto exp(\alpha \times U_s(o, h, C_s))$
  (where α is a normalizing constant)

Listener behavior: the probability $P_l(h|o, C_l)$ that the listener interprets the meaning of o as h in the listener's context model C_l depends on the prior probability $P(h)$ of the speaker having h in mind (e.g. based on certain sociocultural convention) and on the probability $P_s(o|h, C_l)$ that the speaker uses o to convey h in C_l, estimated by the listener.

- $P_l(h|o, C_l) \propto P(h) \times P_s(o|h, C_l)$

Based on this framework, we can develop a minimally workable model that accounts for the emergence of sequence-based social intents in marked linguistic environments where NTUs occur (cf. Acton and Burnett (2019) for social meaning):

- Hidden: the speaker's social intents.
- Observed: Topics chosen / topic transitions.
- Cost: content-based complexity of the topic transitions (e.g. from the perspective of cognitive processing).
- Utility: subtraction of the cost from the coherence measure (which reflects both types of coherence: topical and non-topical).

However, this model design is not robust enough to predict the emergence of the newly classified sequence-based social intents due to the simplicity of the utility function. Specifically, the forthright division of labor between the cost and coherence measure does not capture the real interactions between the components of these metric concepts, such as multiple sociolinguistic dimensions of the discourse context. We will address this challenge in our further work.

8 Conclusion and Future Work

In this paper, we present a pilot annotation study[13] as a first step towards a dialogue model which is capable of rationalizing NTUs and conversational coherence in social talk. Analyzing the interaction between the identified NTUs and their discourse context, we discover a set of patterns of DTCs, represented by the NTUs. Based on these patterns, we propose a simple classification of NTUs in social talk, yet introducing new sequence-based social intents that traditional taxonomies of speech acts do not capture. These intents not only adequately account for non-topical coherence in social talk but also convincingly demonstrate social talk as a sophisticated form of goal-directed rational interactions. We hypothesize that the Bayesian game-theoretic framework, which explicitly models the interactive and rational aspects of social interaction, is a sensible architecture for handling social talk.

Next, we aim to develop an actionable Bayesian game-theoretic model for social talk, focusing on decomposing its utility function. Particularly, we seek to learn from social interaction work such as Stevanovic and Koski (2018) for designing the ***goal-directedness*** aspect of the model.

[12] These intents should be taken with the caveat concerning the cross-cultural generalization about their validity.

[13] The annotation results can be accessible upon the evidence of the possession of SPICE-Ireland corpus.

Acknowledgments

We would like to thank Marine Riou for fully sharing her annotation methodology and data of topic transitions in conversation so that we can gain important insights into her work. We are extremely grateful to Maarten van Gompel, the main author of the FoLiA annotation format and FLAT annotation tool, who provided key technical solutions and support to our annotation project. We owe a debt of gratitude to the organizers and attendees of the inaugural Natural Language, Dialog and Speech (NDS) Symposium, who gave us a unique opportunity to present our work, receive constructive feedback, and experience genuine interest from the audience. We would like to thank the 2020 ACL SRW Pre-submission Mentorship Program for providing valuable suggestions to improve the readability, overall presentation, and technical level of our paper to put it on a par with other accepted submissions. Our deepest gratitude goes to the anonymous reviewers and post-acceptance mentor, Vicky Zayats, whose detailed comments helped us further develop our paper into a fine piece of work that we are very happy with. Last but not least, we are very grateful to all of the readers and look forward to having thought-provoking discussions with you at https://osf.io/nvtkq/, where the live version of our paper is located.

References

Eric K. Acton and Heather Burnett. 2019. Markedness, rationality, and social meaning. Poster session presented at LSA 2019 Annual Meeting, New York, NY.

James Allen. 1995. *Natural Language Understanding*. Pearson.

Nicholas Asher and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press.

Betty J. Birner. 2006. Inferential relations and non-canonical word order. *Drawing the boundaries of meaning: Neo-Gricean studies in pragmatics and semantics in honor of Laurence R. Horn*, pages 31–51.

Heather Burnett. 2019. Signalling games, sociolinguistic variation and the construction of style. *Linguistics and Philosophy*.

Lynn Carlson and Daniel Marcu. 2001. Discourse tagging reference manual. *ISI Technical Report ISI-TR-545*, 54:56.

Rebecca Clift. 2016. *Conversation Analysis*. Cambridge University Press.

John W Du Bois, Wallace L Chafe, Charles Meyer, Sandra A Thompson, and Nii Martey. 2000. Santa barbara corpus of spoken american english. *CD-ROM. Philadelphia: Linguistic Data Consortium*.

Michael C. Frank and Noah D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998.

Michael Franke. 2009. *Signal to act: Game theory in pragmatics*. Ph.D. thesis, Universiteit van Amsterdam.

Emer Gilmartin, Christian Saam, Brendan Spillane, Maria O'Reilly, Ketong Su, Arturo Calvo Devesa, Loredana Cerrato, Killian Levacher, Nick Campbell, and Vincent Wade. 2018. The ADELE corpus of dyadic social text conversations: Dialog act annotation with ISO 24617-2. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, pages 4016–4022, Miyazaki, Japan.

Jonathan Ginzburg. 2012. *The Interactive Stance: Meaning for Conversation*. Oxford University Press.

Maarten van Gompel, Ko van der Sloot, Martin Reynaert, and Antal van den Bosch. 2017. *FoLiA in Practice: The Infrastructure of a Linguistic Annotation Format*, pages 71–82. Ubiquity Press.

H Paul Grice. 1975. Logic and conversation. *Syntax and Semantics*, 3:41–58.

Barbara J Grosz and Candace L Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational linguistics*, 12(3):175–204.

Barbara J Grosz, Scott Weinstein, and Aravind K Joshi. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational linguistics*, 21(2):203–225.

ISO 24617-2. 2012. Language resource management – semantic annotation framework (SemAF) – part 2: Dialogue acts. Technical report, International Organization for Standardization.

ISO 24617-8. 2016. Language resource management – semantic annotation framework (SemAF) – part 8: Semantic relations in discourse, core annotation schema (DR-core). Technical report, International Organization for Standardization.

Gail Jefferson. 1984. On stepwise transition from talk about a trouble to inappropriately next-positioned matters. *Structures of social action: Studies in conversation analysis*, pages 191–222.

Jeffrey L Kallen and John Monfries Kirk. 2012. *SPICE-Ireland: A User's Guide; Documentation to Accompany the SPICE-Ireland Corpus: Systems of Pragmatic Annotation in ICE-Ireland*. Cló Ollscoil na Banríona.

Tina Klüwer. 2011. "I like your shirt" - Dialogue acts for enabling social talk in conversational agents. In *Proceedings of the 11th International Workshop on Intelligent Virtual Agents (IVA 2011)*, pages 14–27, Reykjavik, Iceland. Springer.

Staffan Larsson and David R Traum. 2000. Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural language engineering*, 6(3-4):323–340.

David Lewis. 1969/2002. *Convention: A philosophical study*. John Wiley & Sons.

Ellen F Prince. 1992. The ZPG letter: Subjects, definiteness, and information-status. *Discourse description: Diverse analyses of a fund raising text*, pages 295–325.

Ines Rehbein, Merel Scholman, and Vera Demberg. 2016. Annotating discourse relations in spoken language: A comparison of the PDTB and CCR frameworks. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, pages 1039–1046, Portorož, Slovenia. European Language Resources Association (ELRA).

Rachel Reichman. 1978. Conversational coherency. *Cognitive science*, 2(4):283–327.

Marine Riou. 2015. *The grammar of topic transition in American English conversation. Topic transition design and management in typical and atypical conversations (schizophrenia)*. Ph.D. thesis, Université Sorbonne Paris Cité.

Craige Roberts. 1996/2012. Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5:6–1.

Ted JM Sanders, Wilbert PM Spooren, and Leo GM Noordman. 1992. Toward a taxonomy of coherence relations. *Discourse processes*, 15(1):1–35.

Remko JH Scha, Bertram C Bruce, and Livia Polanyi. 1986. Discourse understanding. *Center for the Study of Reading Technical Report; no. 391*.

John R. Searle. 1976. A classification of illocutionary acts. *Language in Society*, 5(1):1–23.

Melisa Stevanovic and Sonja E Koski. 2018. Intersubjectivity and the domains of social interaction: Proposal of a cross-sectional approach. *Psychology of Language and Communication*, 22(1):39–70.

Bonnie Webber, Rashmi Prasad, Alan Lee, and Aravind Joshi. 2016. A discourse-annotated corpus of conjoined VPs. In *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016)*, pages 22–31, Berlin, Germany. Association for Computational Linguistics.

Bonnie Webber, Rashmi Prasad, Alan Lee, and Aravind Joshi. 2019. The Penn Discourse Treebank 3.0 annotation manual. Technical report, University of Edinburgh.

JM van der Zwaan, V Dignum, and CM Jonker. 2012. A BDI dialogue agent for social support: Specification and evaluation method. In *AAMAS 2012: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems, Workshop on Emotional and Empathic Agents, Valencia, Spain; authors' version*. International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS).

A Coherence relation annotation in practice

As the input data of this annotation task includes different useful information layers, namely the PDTB 3.0 discourse relations of Disco-SPICE and pragmatic annotation of SPICE-Ireland, the FoLiA format is selected for data representation because this rich XML-based annotation format accommodates multiple linguistic annotation types with arbitrary tagsets and is accompanied by FLAT, a modern web-based annotation tool whose user-interface can show different linguistic annotation layers at the same time (van Gompel et al., 2017). Specifically, each dialogue is a sequence of utterances, as shown in Figure 1, each of which includes:

- the 'speaker' token (highlighted in green), combining the dialogue ID and the speaker ID, whose "Description" field contains SPICE-Ireland pragmatic annotations (see Figure 3 for an example of an utterance annotated as a directive, i.e. *<dir>*, and a complete intonational unit, i.e. ended with %, whose final token *them* is spoken in a rising tone, i.e. *2*),
- the tokenized content, which may consist of:
  - explicit discourse connectives or AltLex expressions, i.e. non-connective expressions which lexicalize the corresponding discourse relations, (highlighted in various colors).
  - implicit discourse connective tokens (in gray).
  - real *[None]* tokens (in black), equivalent to empty event tokens in the original Disco-SPICE *.xml* file.
  - hidden *[None]* tokens (in gray), placeholders of EntRel discourse relations.

Figure 2 shows that when a token is hovered over, it is highlighted in black while its text turns yellow, and its annotation layers are displayed in a pop-up box.

Figure 3 shows that when a token is clicked, it is highlighted in yellow, and its annotation layers become editable in the **Annotation Editor**.

The annotation of one coherence relation is treated as the annotation of one 'connective' entity and two 'argument' chunks. Each 'connective' entity has its co-index with its 'argument' chunks in its "Description" field. Figure 4 shows that the 'connective' entity *in_particular* has its co-index *72* with its 'argument' chunks, namely *ARG1-72* and *ARG2-72*. This is an example of an implicit relation inherited from Disco-SPICE.

Figures 5, 6 and 7 show several newly annotated relations, namely *propositionalQ*, *EntRel*, and *feedback* respectively. Notice that the 'argument' chunks only need associating with the 'speaker' tokens of the utterances containing the actual chunks. To annotate a 'connective' entity that does not connect to any real text token, we create a hidden token *[None]* right before the 'speaker' token of the '2nd argument' chunk in the corresponding relation.

B Examples of DTCs

Table 3 displays the DTCs, corresponding to the NTUs of the excerpt shown in Table 1. ICP and OCP stand for initiating conversational participant and other conversational participant(s) respectively (Grosz and Sidner, 1986).

Figure 1: FLAT-based representation of the excerpt shown in Table 1.

Utt.	Preceding topic	Current topic	Involved CPs	Topic change type
119	Jamie's husband having another woman	Reaction to an event in the utterance situation - Discourse New	ICP (A) as the speaker	Grosz and Sidner's true interruption
123	An event happening in ICP's place	New focused entity: OCP's day - Discourse New	OCP (B) as the hearer	Hearer-centric related topic

Table 3: Examples of DTC patterns in the excerpt shown in Table 1.

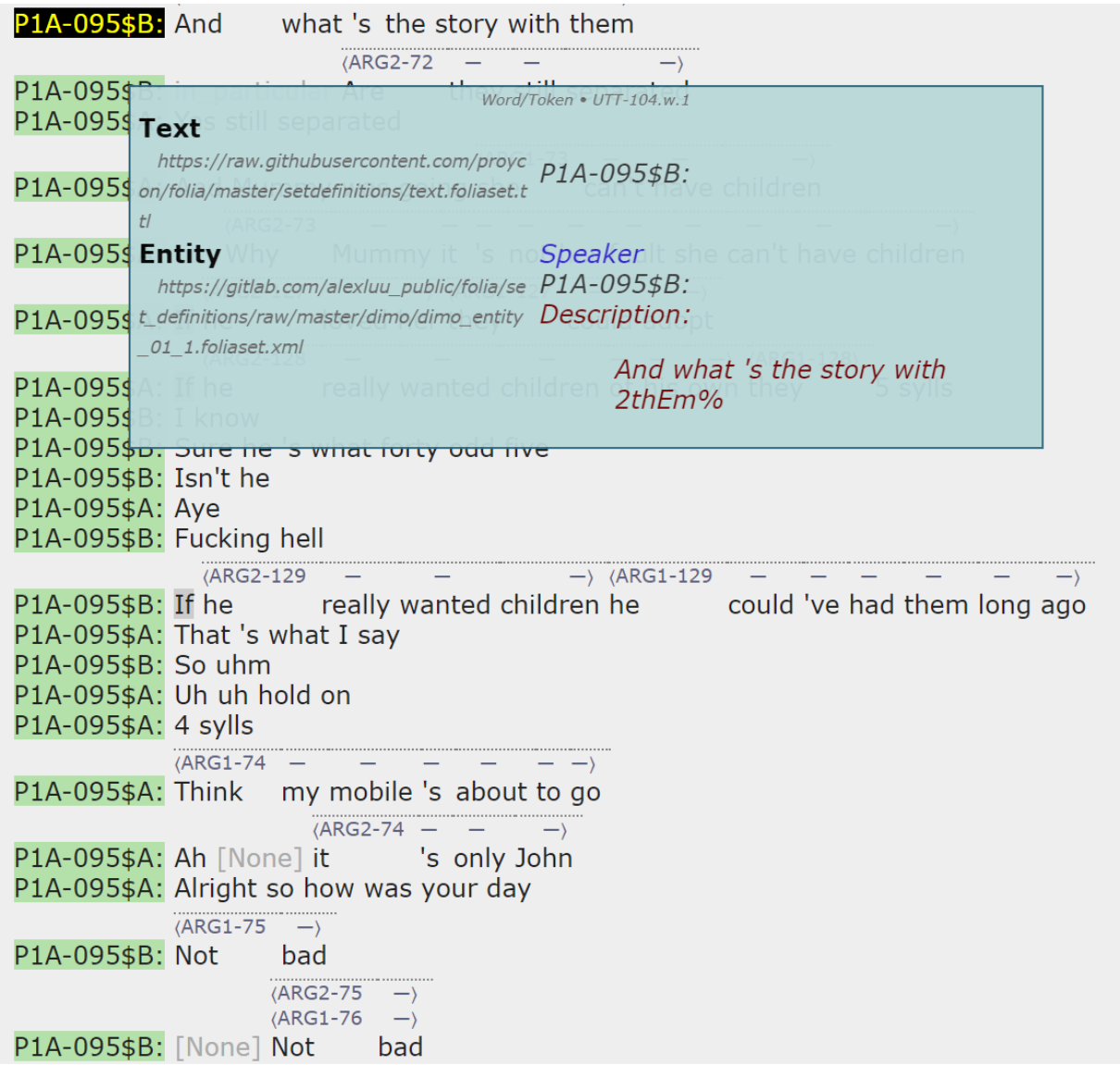

Figure 2: Quick access to the annotation of a token in FLAT.

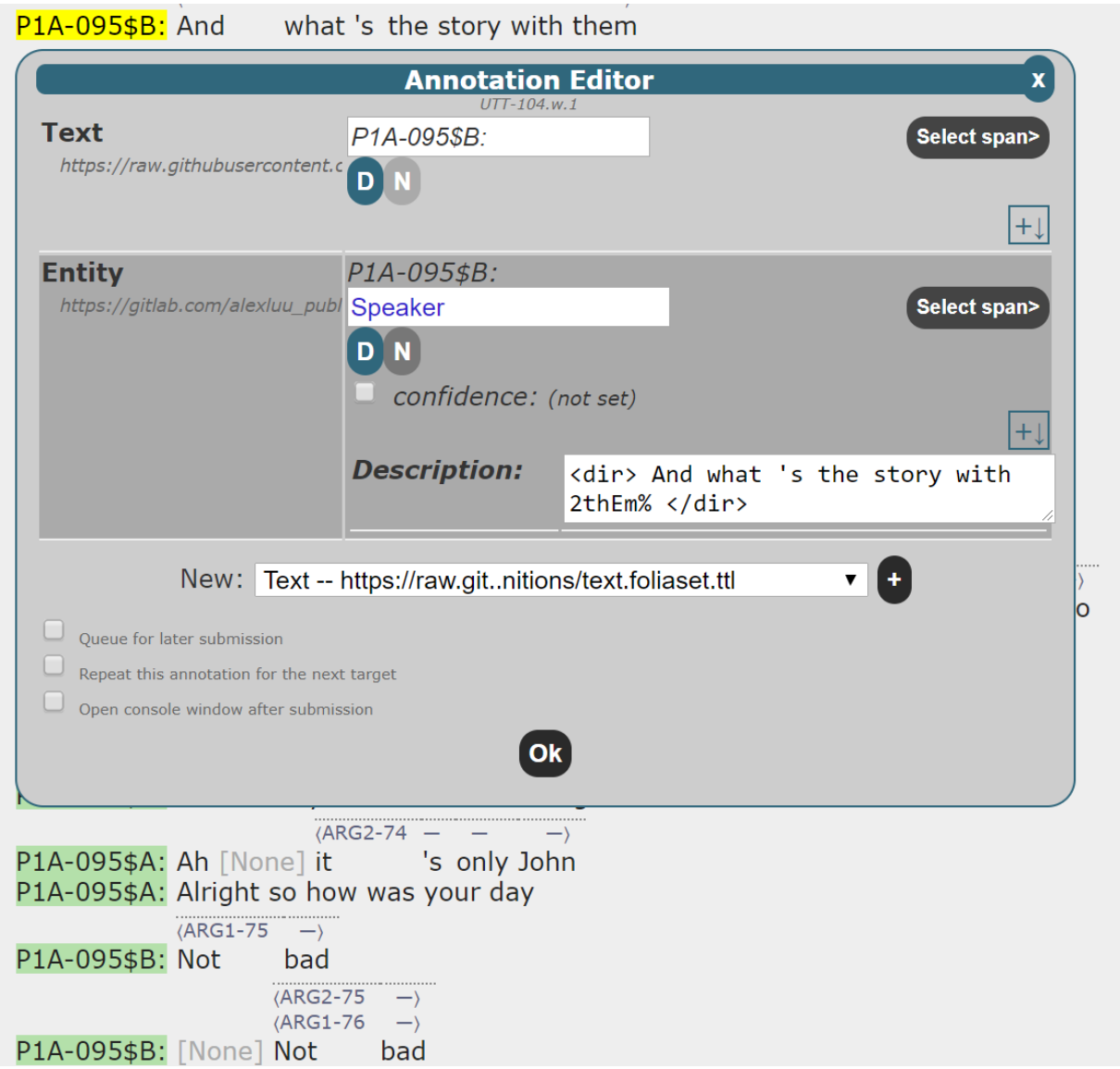

Figure 3: **Annotation Editor** for a token in FLAT.

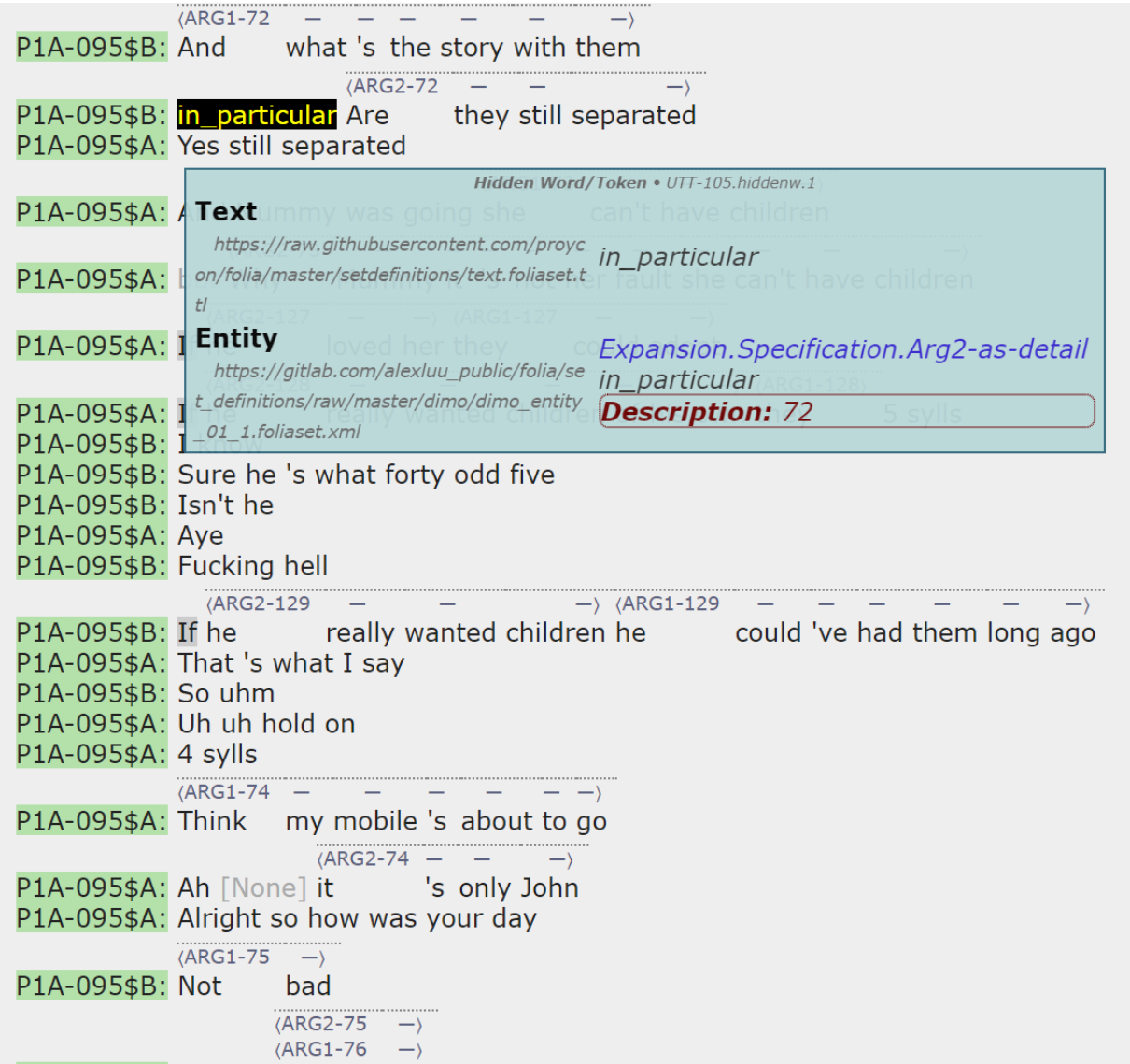

Figure 4: FLAT-based representation of a coherence relation inherited from Disco-SPICE.

Figure 5: FLAT-based representation of a *propositionalQ* relation.

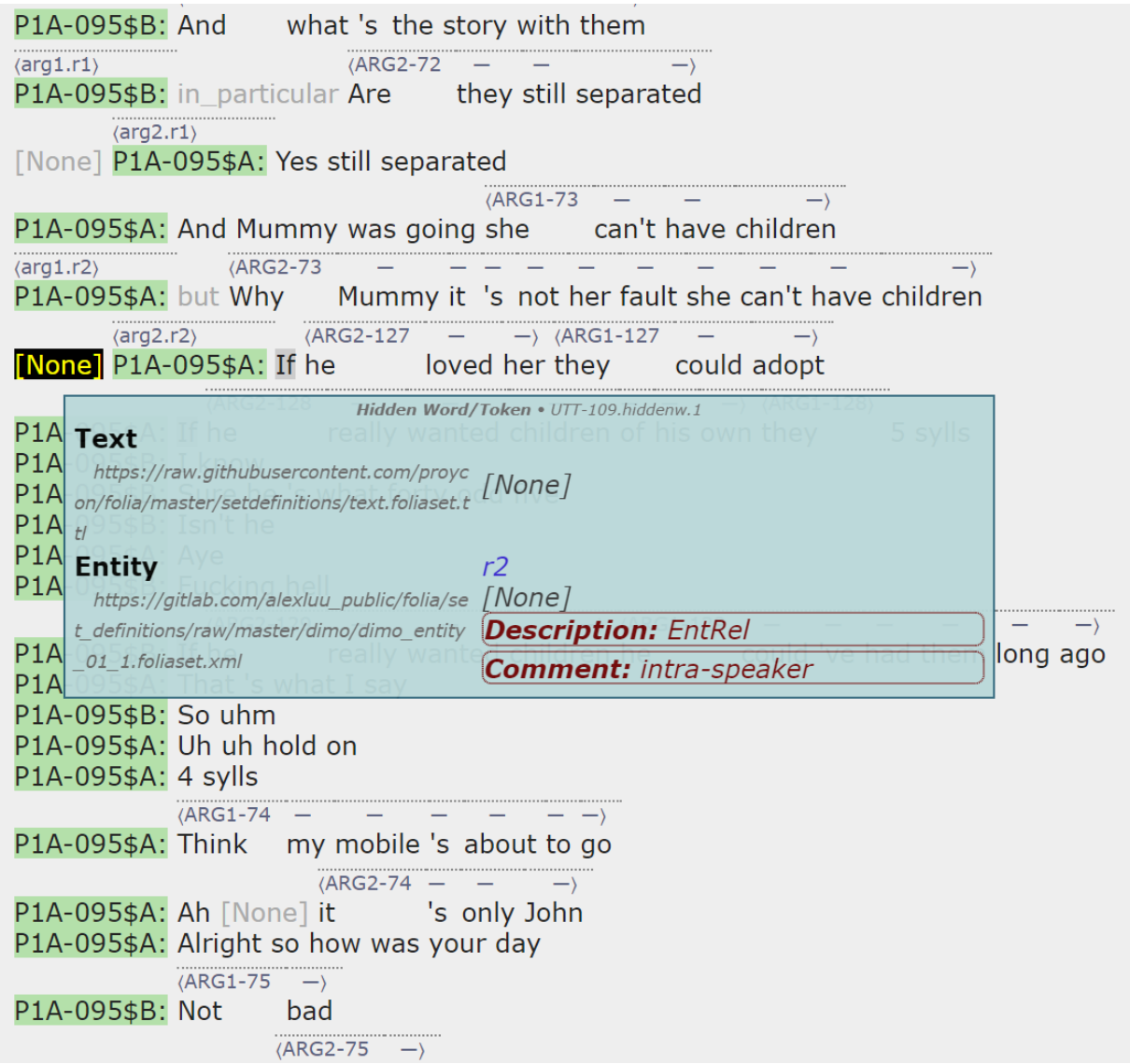

Figure 6: FLAT-based representation of an *EntRel* relation.

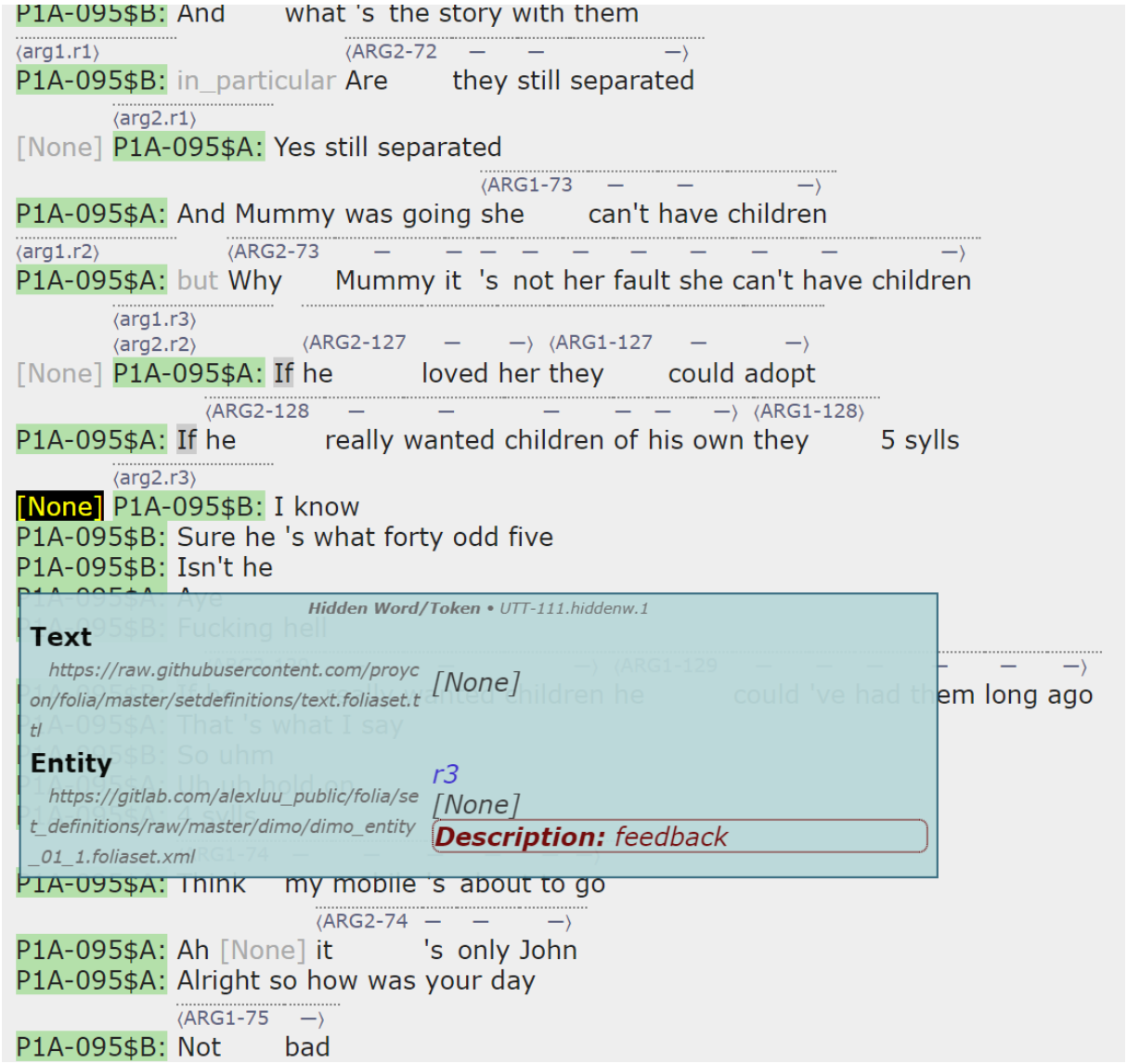

Figure 7: FLAT-based representation of a *feedback* relation.