Anticipating the Future: Forward-Looking Discourse Based on Question Under Discussion (QUD)

Abstract. This paper presents an integrated approach that combines Roberts (1996)' stack QUD with the idea of potential question (Onea 2013) to create a dynamic, forward-looking and answer-agnostic QUD generation and updating system. Unlike previous QUD modeling approaches that focused on discourse recovery, our model emphasizes the flow of questions and answers within the discourse. We demonstrate that our model captures distinct discourse relations compared to the recovery model.

1. Introduction. Reproducing human-like discourse in machines presents a formidable challenge. Previous attempts (e.g.,Mann & Thompson 1988; Carlson & Marcu 2001) take advantage of a "complete" piece of discourse, allowing for higher contextual resolution by leveraging preceding and following context. This approach carries the potential risk of overlooking the dynamic nature of discourse. In an attempt to maintain the dynamicity, this study adopts the linguistic framework of Question Under Discussion (QUD) (Roberts 1996, 2012; Ginzburg et al. 1996; Van Kuppevelt 1995, 1996; Büring 2003). This framework offers a dynamic perspective on discourse moves and has the potential to provide novel insights into discourse relations and coherence (Onea 2013, 2016; Hunter & Abrusán 2015). We hope our approach to computationalizing QUDs to model discourse will provide insights into downstream tasks such as text-planning (Huot et al. 2023), and simplification (Ko et al. 2023; Wu et al. 2023).

2. QUD theory. Our model is grounded in the QUD theory, drawing insights from two foundational works: the stack model originally proposed by Roberts (1996) and the potential question theory introduced by Onea (2013). In Roberts (1996), the stack model operates with specific rules: (1) Overtly asked questions are treated as QUDs and are placed at the topmost position of a stack. (2) Discourse progresses by introducing sub-questions or providing answers through assertions. (3) Once a question is answered, it is removed from the stack. In addition, Onea (2013) highlights the crucial role of assertions that also influence future questions, that is, QUDs should be one of the potential questions invoked by the assertions. Hence, based on this idea, we establish a stack of potential questions before constructing the actual QUD stack.

Our model demonstrates a two-stage process for predicting the QUD at a specific utterance. First, a stack of potential questions is generated (Generation stage or the Speaker). Then, the continuation of the discourse is juxtaposed with the potential questions. If the continuation is an assertion, the model checks which question is answered by the assertion. On the other hand, if the continuation is a question, the overt question directly serves as the QUD (Confirmation stage or the Addressee).

3. Computationalizing QUDs: the Creation Model. The current work aims to model QUDs in a forward-looking and answer-agnostic manner and we will call this model the "creation model" in contrast to the "recovery model" (e.g., Ko et al. 2023), because in the creation model, potential QUDs are "created" first, and the answer is utilized to confirm the current QUD while in the recovery model, the answer is directly employed to "generate/recover" QUDs. The creation model can be broadly divided into two components: a potential question generator and a QUD confirmer. We add a potential question ranker to ensure the quality of the generated questions.

Question Generator (QG) Our QG is largely based on the QG model presented in Ko et al. (2020), with some adjustments to fit our system. We incorporated a pretrained GPT-2 model (Radford et al. 2019) and fine-tuned it using Ko's INQUISITIVE dataset. In addition, we further rank these generated questions using the random forest classifier in Schricker & Scheffler (2019), which is trained based on features of potential questions (e.g., indefinites invoke potential questions).

QUD Confirmer We opted for BERT (Devlin et al. 2019) as our confirmer, which has demonstrated strong performance in QA and sequence classification tasks. We fine-tuned BERT on a dataset combining DCQA (Ko et al. 2021), and INQUISITIVE-EXTEND (Wadhwa et al. 2023), which provides non-factoid answer and question pairs present in the same texts.

We incorporated these modules into one pipeline as shown in Figure 1. Each question will be indexed in terms of its generation position (index 1) and its corresponding position where it is answered (index 2) in order to track the flow of questions within the discourse. The QUD stack from the previous step will be retained or merged with newly generated PQ stack in case that the QUD we generated can be answered more than once or some QUDs are generated earlier in the discourse but answered relatively late.

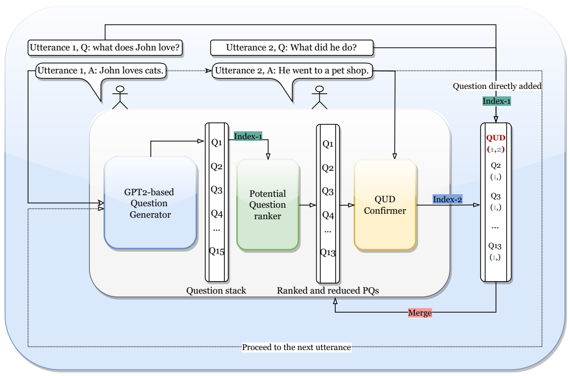

Figure 1. Pipeline for QUD generation and updates (the Creation Model)

4. Evaluation and Implications. We compared the the creation model with the recovery model, in terms of the discourse relations predicted, using pieces from RST treebank Carlson & Marcu (2001). The creation model tends to capture the structure at lower levels on the RST tree, whereas the recovery model focuses more on high-level relations (see Figure 2 as one example). We also conducted linguistic analysis in terms of ellipsis recoverability (Weir et al. 2017; Griffiths 2019), and alternative elimination (Coppock & Beaver 2014; Warstadt 2020). The results show (1) the current model builds connections between elided sites and their antecedents (that is, some syntactic understanding), and (2) it fails to detect the role of the exclusives (e.g., finer-grained lexical semantics of *just* and *only*) that eliminate unnecessary potential questions.

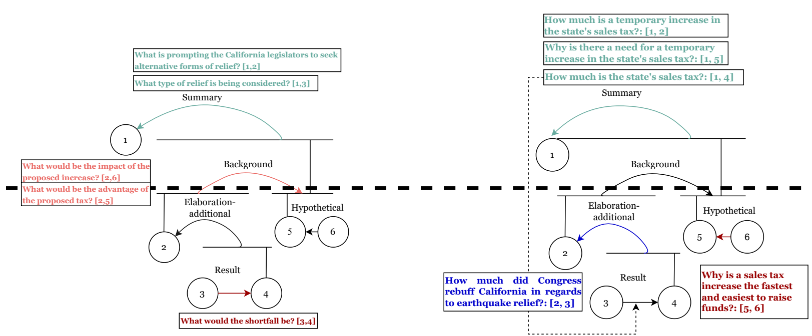

Figure 2. Comparing the creation model and the recovery model based on RST

References

Büring, Daniel. 2003. On d-trees, beans, and b-accents. *Linguistics and philosophy* 26. 511–545.

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

Coppock, Elizabeth & David I Beaver. 2014. Principles of the exclusive muddle. *Journal of Semantics* 31(3). 371–432.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee & Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. *ArXiv* abs/1810.04805. https://api.semanticscholar.org/CorpusID:52967399.

Ginzburg, Jonathan et al. 1996. Dynamics and the semantics of dialogue. *Logic, language and computation* 1. 221–237.

Griffiths, James. 2019. A q-based approach to clausal ellipsis: Deriving the preposition stranding and island sensitivity generalisations without movement. *Glossa: a journal of general linguistics* 4(1).

Hunter, Julie & Marta Abrusán. 2015. Rhetorical structure and quds. In *Jsai international symposium on artificial intelligence*, 41–57. Springer.

Huot, Fantine, Joshua Maynez, Shashi Narayan, Reinald Kim Amplayo, Kuzman Ganchev, Annie Louis, Anders Sandholm, Dipanjan Das & Mirella Lapata. 2023. Text-blueprint: An interactive platform for plan-based conditional generation. *arXiv preprint arXiv:2305.00034* .

Ko, Wei-Jen, Tengyang Chen, Yiyan Huang, Greg Durrett & Junyi Jessy Li. 2020. Inquisitive question generation for high level text comprehension. In *Conference on empirical methods in natural language processing*, .

Ko, Wei-Jen, Cutter Dalton, Mark P. Simmons, Eliza Fisher, Greg Durrett & Junyi Jessy Li. 2021. Discourse comprehension: A question answering framework to represent sentence connections. In *Conference on empirical methods in natural language processing*, .

Ko, Wei-Jen, Yating Wu, Cutter Dalton, Dananjay Srinivas, Greg Durrett & Junyi Jessy Li. 2023. Discourse analysis via questions and answers: Parsing dependency structures of questions under discussion. In *Findings of the association for computational linguistics: Acl 2023*, 11181–11195. Toronto, Canada: Association for Computational Linguistics. https://aclanthology.org/2023.findings-acl.710.

Mann, William C & Sandra A Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text-interdisciplinary Journal for the Study of Discourse* 8(3). 243–281.

Onea, Edgar. 2013. Potential questions in discourse and grammar. *University of Göttingen Habilitation thesis* 10.

Onea, Edgar. 2016. *Potential questions at the semantics-pragmatics interface*, vol. 33. Brill.

Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1(8). 9.

Roberts, Craige. 1996. Information structure in discourse: Toward a unified theory of formal pragmatics. *Ohio State University Working Papers in Linguistics* 49. 91–136.

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

Schricker, Luise & Tatjana Scheffler. 2019. Ranking of potential questions. In *Proceedings of the 57th annual meeting of the association for computational linguistics: Student research workshop*, 143–148.

Van Kuppevelt, Jan. 1995. Discourse structure, topicality and questioning. *Journal of linguistics* 31(1). 109–147.

Van Kuppevelt, Jan. 1996. Directionality in discourse: Prominence differences in subordination relations1. *Journal of semantics* 13(4). 363–395.

Wadhwa, Manya, Jifan Chen, Junyi Jessy Li & Greg Durrett. 2023. Using natural language explanations to rescale human judgments.

Warstadt, Alex. 2020. "just" don't ask: Exclusives and potential questions. In *Proceedings of sinn und bedeutung*, vol. 24 2, 373–390.

Weir, Andrew, Robert Truswell, Chris Cummins, Caroline Heycock, Brian Rabern & Hannah Rohde. 2017. Cointensional questions and their implications for fragment answers. In *Proceedings of sinn und bedeutung*, vol. 21, 1290–1306.

Wu, Yating, William Sheffield, Kyle Mahowald & Junyi Jessy Li. 2023. Elaborative simplification as implicit questions under discussion. *arXiv preprint arXiv:2305.10387* .