

# A Discourse-based Approach to Generating *Why*-Questions from Texts

**Rashmi Prasad**

University of Pennsylvania  
rjprasad@seas.upenn.edu

**Aravind Joshi**

University of Pennsylvania  
joshi@seas.upenn.edu

## Abstract

We address the subtask of generating *why*-questions from texts and propose the use of causal relations annotated in the Penn Discourse TreeBank for evaluating content-selection methods for *why*-question generation. Our initial experiments show that 71% of an independently developed data set of *why*-questions can be correlated with causal relations annotated in the PDTB.

## 1 Introduction

Within the question generation (Rus et al., 2007) task, we address the subtask of generating *why*-questions (*why*-QG) from complete documents, specifically those *why*-questions that are “supported” by the text, and that crucially do not make use of world knowledge extraneous to the text. As for other types of questions, there are two aspects to the *why*-QG task that must be separated - identifying the source of the questions, i.e., *content selection*, and generating the questions from the identified sources, i.e., *planning and realization*. In this paper, we focus on the content selection aspect of *why*-QG, and propose the use of a recently released data resource, the Penn Discourse Treebank (PDTB) (Prasad et al., 2008), as a testbed for identifying the source of *why*-questions.<sup>1</sup> *Why*-questions typically have their source in causal relations (e.g., reason, result, purpose, or justification relations) expressed

between events, states, or propositions mentioned in the text, and these relations comprise a subset of the discourse relations annotated in the PDTB. Thus, given an input document from the PDTB corpus, content selection methods for *why*-QG - essentially, methods for automatic identification of causal relations - can be evaluated by standard *recall* and *precision* measures through reference to PDTB causal relations annotated in the same document.

To evaluate the feasibility of using the PDTB discourse relations for the *why*-QG task, we examined the correlation between an independently developed data-set of *why*-questions and answers (Verberne et al., 2007) created over a subset of the PDTB source corpus, and the PDTB relations over the same subset. Our initial findings show that 71% of the QA pairs in the *why*-QA data have their source in causal relations annotated in the PDTB, thus suggesting that the PDTB could be useful for evaluation in *why*-QG. In the remainder of this paper, we present our analysis of the *why*-QA data set with respect to the PDTB, and briefly discuss some issues raised by the QA pairs unaccounted for in the PDTB.

## 2 Using the PDTB for Evaluating Content Selection in *why*-QG

The PDTB contains annotations of “explicit” and “implicit” discourse relations and their (two) arguments (annotated as Arg1 and Arg2), their (possibly multiple) senses, and their attributions on the WSJ corpus (Marcus et al., 1993). Discourse relations, such as causal, contrastive, temporal, conjunctive, and elaboration relations, are taken to hold between the “abstract objects” mentioned in the text,

<sup>1</sup>The PDTB (<http://www.seas.upenn.edu/~pdtb>) is available through the Linguistic Data Consortium (<http://www ldc.upenn.edu>).

such as events, states, facts, and propositions. There are two main reasons for why we believe the PDTB may be particularly conducive to content selection in *why*-QG. First, the PDTB is theory-independent, in that it is not committed to building any particular type of structure over the entire text (e.g., a tree structure), so that “relation discovery” is not unduly constrained by an assumed data structure. Second, PDTB discourse relations are linked to lexical expressions when they are explicit in the text (Examples 1 and 2), thus facilitating their automatic identification.<sup>2</sup> And while implicit (inference-dependent) discourse relations are annotated, they are marked only between adjacent sentences (Ex. 3) via the insertion of an “implicit” connective to express the inferred relation, and are thus constrained inferences.

- (1) ... *increased carbon dioxide emissions*, ..., *will cause the Earth to warm up* because **carbon dioxide prevents heat from escaping into space.** (cause:reason)<sup>3</sup>
- (2) ... *its field staff signed up the 500 schools in 238 school districts after only eight weeks* and **company executives now expect to reach their start-up goal of 1,000 schools before the end of this year.** (expansion:conjunction/cause:result)
- (3) ..., *researchers analyzed the changes in concentrations of two forms of oxygen.* Implicit=because **These measurements can indicate temperature changes**, ... (cause:reason)

In its current state, the PDTB contains fewer relations than could have been identified given its approach, since some practical decisions were made to not annotate certain types of relations. For example, intra-sentential implicit relations between clauses and implicit relations across paragraphs were not annotated. While these gaps in the annotation can be filled eventually by models trained on the existing annotations, we decided to manually extend the PDTB annotations of 3 texts for our analysis below.<sup>4</sup>

<sup>2</sup>PDTB discourse relations can be lexicalized as connectives drawn from well-defined syntactic classes (subordinating/coordinating conjunctions and adverbials), but also as non-connective expressions, called *alternative lexicalizations* (“AltLex”).

<sup>3</sup>In the PDTB examples, relations are underlined; Arg1 appears in italics; Arg2 is in boldface; senses appear in parentheses; multiple senses are separated by “/”.

<sup>4</sup>We extended the annotations for only 3 texts due to time constraints. However, we plan to extend a larger subset of the

## 2.1 Evaluating the PDTB Relations with a *Why*-QA Data Set

With the recent emergence of interest in *why*-question answering (*why*-QA) has also come some intensive development of data-collection for *why*-questions (and answers), in particular by Verberne et al. (2007). For our study, we used Verberne et al.’s QA pairs created from 7 texts in the WSJ corpus, since this is a subset of the same corpus over which the PDTB has been annotated. Verberne et al set up an elicitation experiment wherein native speakers were asked to read the texts, formulate *why*-questions for which the answer could be found in the text (i.e., were “supported” by the text), and also formulate the answers to each of their questions. Each text was given to multiple subjects so that there is much overlap in the topics of the questions (roughly 3 to 4 questions per topic).<sup>5</sup> Given that the task of the humans was to generate as many *why*-questions as they could after having read the text, this data can be taken to be exhaustively representative of all the *why*-questions that could be supported by the text.

For this paper, we examined the QA pairs for 3 of the 7 texts in Verberne’s collection, for which we had extended the PDTB annotations. Out of a total of 177 questions from these 3 texts, we found we had to exclude several questions, for the following reasons. First, we excluded 30 questions since they (or their answers) were not supported by the text.<sup>6</sup> Second, because of the considerable overlap between the topics of questions, we only selected one QA pair within each topic group.<sup>7</sup> This selection was done randomly. After these exclusions and groupings, the final set of QA pairs we obtained for the 3 texts was 38. We manually examined these 38 questions to see if they were associated with the arguments of a PDTB relation in the relevant text, also noting the sense of the relation. Our analysis showed

corpus and make it available for community-wide use.

<sup>5</sup>Since a single question topic has multiple realizations in this data set, it is also very useful in principle for the *planning and realization* aspect of *why*-QG.

<sup>6</sup>For a question-answering application, recognition of causal relations can also be useful for identifying unsupported *why*-questions, and allow the system to report this to the user.

<sup>7</sup>Data on unsupported questions (and answers) and topic groupings of the questions were not part of the public distribution of the *why*-QA data, but were obtained separately from Suzan Verberne upon request.

the following. On the one hand, all the causal relations annotated in the PDTB were associated with a question in the *why*-QA data, providing evidence for the exhaustiveness of the *why*-QA data. On the other hand, 28 of the 38 QA pairs were associated with a PDTB relation, 27 of which were causal relations. (The single non-causal relation was specification (a type of elaboration).) From the causal relations in Examples (1), (2), and (3) above, taken from the 3 texts we studied, the following *why*-questions were generated in the *why*-QA data.

- (4) (For Ex. 1): Why will increased carbon dioxide emissions cause the Earth to heat up?
- (5) (For Ex. 2): Why does Channel One expect to reach their start-up goal of 1000 schools?
- (6) (For Ex. 3): Why did researchers analyze the changes in concentration of two forms of oxygen?

Thus, the PDTB causal relations can be seen as providing the source for 71% (27/38) of the *why*-questions. The remaining questions had one of the following problems: (a) the question and answer were in distinct clauses/sentences unrelated to each other by a PDTB relation (16% of all questions); (b) the question and answer appeared within a single clause, with the relation expressed with a causal verb/preposition (11% of all questions). For the question *Why is Channel One controversial?* and its answer *Four 30-second TV ads during each show have drawn protests from educators*, Example (7) shows the source fragments (in italics). These two fragments are unrelated in the PDTB because they are 8 sentences apart, requiring a complex inference, such as “if a show draws protests, then it must be controversial”. We plan to explore whether some of these cases can be identified by computing long-distance causal relations over chains of relations.

- (7) Whittle Communications L.P., . . . , said it has signed 500 schools in 24 states to subscribe to the controversial Channel One news program . . . . . (8 sentences) . . .  
Subscribing schools get the 12-minute daily Channel One news program, *whose four 30-second TV ads during each show have drawn protests from educators*.

The answer fragment in Example (7) is also the source for another QA pair, with the question as

*Why have educators protested Channel One?* and the answer as *Because it shows ads to pupils*. Here, the causal relation is unannotated in the PDTB because it is expressed intra-clausally through the potentially causal verb *drawn*. However, we believe that it should be possible to combine the PDTB data with independent resources that attempt to annotate intra-clausal relations.

### 3 Conclusion

In this paper, we proposed using an existing resource of annotated discourse relations, the Penn Discourse TreeBank (PDTB), for evaluating approaches to content selection for *why*-question generation. We examined the coverage of PDTB causal relations over an independently developed data set of *why*-questions and answers. Our initial findings showed a 71% correlation between the two data sets. We conclude that the PDTB could be useful for evaluating content selection methods in *why*-QG.

### Acknowledgements

We are grateful to Suzan Verberne for freely providing the *why*-QA data. We also thank Alan Lee, Bonnie Webber, and the anonymous reviewers for their helpful comments. This work was partially supported by NSF grants EIA-02-24417, EIA-05-63063, and IIS-07-05671.

### References

- Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of english: The Penn Treebank. *Computational Linguistics*, 19(2):313–330.
- Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltasakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings of LREC*.
- Vasile Rus, Zhiqiang Cai, and Arthur C. Graesser. 2007. Evaluation in natural language generation: The question generation task. In *Proceedings of the Workshop on Shared Tasks and Comparative Evaluation in Natural Language Generation*, pages 20–21.
- Suzan Verberne, Lou Boves, Peter-Arno Coppen, and Nelleke Ostdijk. 2007. Discourse-based answering of *why*-questions. *Traitement Automatique des Langues, Special Issue on Computational Approaches to Document and Discourse*, 47(2):21–41.