

Data-oriented Monologue-to-Dialogue Generation

Paul Piwek

Centre for Research in Computing
The Open University
Walton Hall, Milton Keynes, UK
p.piwek@open.ac.uk

Svetlana Stoyanchev

Centre for Research in Computing
The Open University
Walton Hall, Milton Keynes, UK
s.stoyanchev@open.ac.uk

Abstract

This short paper introduces an implemented and evaluated monolingual Text-to-Text generation system. The system takes monologue and transforms it to two-participant dialogue. After briefly motivating the task of monologue-to-dialogue generation, we describe the system and present an evaluation in terms of fluency and accuracy.

1 Introduction

Several empirical studies show that delivering information in the form of a dialogue, as opposed to monologue, can be particularly effective for education (Craig et al., 2000; Lee et al., 1998) and persuasion (Suzuki and Yamada, 2004). Information-delivering or expository dialogue was already employed by Plato to communicate his philosophy. It is used primarily to convey information and possibly also make an argument; this in contrast with dramatic dialogue which focuses on character development and narrative.

Expository dialogue lends itself well for presentation through computer-animated agents (Prendinger and Ishizuka, 2004). Most information is however locked up as text in leaflets, books, newspapers, etc. Automatic generation of dialogue from text in monologue makes it possible to convert information into dialogue as and when needed.

This paper describes the first *data-oriented* monologue-to-dialogue generation system which relies on the automatic mapping of the discourse relations underlying monologue to appropriate se-

quences of dialogue acts. The approach is data-oriented in that the mapping rules have been automatically derived from an annotated parallel monologue/dialogue corpus, rather than being hand-crafted.

The paper proceeds as follows. Section 2 reviews existing approaches to dialogue generation. Section 3 describes the current approach. We provide an evaluation in Section 4. Finally, Section 5 describes our conclusions and plans for further research.

2 Related Work

For the past decade, generation of information-delivering dialogues has been approached primarily as an AI planning task. André et al. (2000) describe a system, based on a centralised dialogue planner, that creates dialogues between a virtual car buyer and seller from a database; this approach has been extended by van Deemter et al. (2008). Others have used (semi-) autonomous agents for dialogue generation (Cavazza and Charles, 2005; Mateas and Stern, 2005).

More recently, first steps have been taken towards treating dialogue generation as an instance of Text-to-Text generation (Rus et al., 2007). In particular, the T2D system (Piwek et al., 2007) employs rules that map text annotated with discourse structures, along the lines of Rhetorical Structure Theory (Mann and Thompson, 1988), to specific dialogue sequences. Common to all the approaches discussed so far has been the manual creation of generation resources, whether it be mappings from knowledge representations or discourse to dialogue structure.

With the creation of the publicly available¹ CODA parallel corpus of monologue and dialogue (Stoyanchev and Piwek, 2010a), it has, however, become possible to adopt a data-oriented approach. This corpus consists of approximately 700 turns of dialogue, by acclaimed authors such as Mark Twain, that are aligned with monologue that was written on the basis of the dialogue, with the specific aim to express the same information as the dialogue.² The monologue side has been annotated with discourse relations, using an adaptation of the annotation guidelines of Carlson and Marcu (2001), whereas the dialogue side has been marked up with dialogue acts, using tags inspired by the schemes of Bunt (2000), Carletta et al. (1997) and Core and Allen (1997). As we will describe in the next section, our approach uses the CODA corpus to extract mappings from monologue to dialogue.

3 Monologue-to-Dialogue Generation Approach

Our approach is based on five principal steps:

- I *Discourse parsing*: analysis of the input monologue in terms of the underlying discourse relations.
- II *Relation conversion*: mapping of text annotated with discourse relations to a sequence of dialogue acts, with segments of the input text assigned to corresponding dialogue acts.
- III *Verbalisation*: verbal realisation of dialogue acts based on the dialogue act type and text of the corresponding monologue segment.
- IV *Combination* Putting the verbalised dialogues acts together to create a complete dialogue, and
- V *Presentation*: Rendering of the dialogue (this can range for simple textual dialogue scripts to computer-animated spoken dialogue).

¹computing.open.ac.uk/coda/data.html

²Consequently, the corpus was not constructed entirely of pre-existing text; some of the text was authored as part of the corpus construction. One could therefore argue, as one of the reviewers for this paper did, that the approach is not entirely data-driven, if data-driven is interpreted as ‘generated from unadulterated, free text, without any human intervention needed’.

For step I we rely on human annotation or existing discourse parsers such as DAS (Le and Aibeysinghe, 2003) and HILDA (duVerle and Prendinger, 2009). For the current study, the final step, V, consists simply of verbatim presentation of the dialogue text. The focus of the current paper is with steps II and III (with combination, step IV, beyond the scope of the current paper). Step II is data-oriented in that we have extracted mappings from discourse relation occurrences in the corpus to corresponding dialogue act sequences, following the approach described in Piwek and Stoyanchev (2010). Stoyanchev and Piwek (2010b) observed in the CODA corpus a great variety of Dialogue Act (DA) sequences that could be used in step II, however in the current version of the system we selected a representative set of the most frequent DA sequences for the five most common discourse relations in the corpus. Table 1 shows the mapping from text with a discourse relations to dialogue act sequences (i indicates implemented mappings).

DA sequence	A	C D	C T	E R	M M	TR T
YNQ; Expl	i	i				d
YNQ; Yes; Expl		i	i	i		d
Expl; CmplQ; Expl					i	d
ComplQ; Expl	i/t	i/t		i	i	c
Expl; YNQ; Yes					i	d
Expl; Contrad.			i			d
FactQ; FactA; Expl				i		c
Expl; Agr; Expl				i		d
Expl; Fact; Expl	t					c

Table 1: Mappings from discourse relations (A = Attribution, CD = Condition, CT = Contrast, ER = Explanation-Reason, MM = Manner-Means) to dialogue act sequences (explained below) together with the type of verbalisation transformation TR being d(irect) or c(omplex).

For comparison, the table also shows the much less varied mappings implemented by the T2D system (indicated with t). Note that the actual mappings of the T2D system are directly from discourse relation to dialogue text. The dialogue acts are not explicitly represented by the system, in contrast with the current two stage approach which distinguishes between relation conversion and verbalisation.

Verbalisation, step III, takes a dialogue act type and the specification of its semantic content as given by the input monologue text. Mapping this to the appropriate dialogue act requires mappings that vary in complexity.

For example, *Expl(ain)* can be generated by simply copying a monologue segment to dialogue utterance. The dialogue acts *Yes* and *Agreement* can be generated using canned text, such as “That is true” and “I agree with you”.

In contrast, *ComplQ* (*Complex Question*), *FactQ* (*Factoid Question*), *FactA* (*Factoid Answer*) and *YNQ* (*Yes/No Question*) all require syntactic manipulation. To generate *YNQ* and *FactQ*, we use the CMU Question Generation tool (Heilman and Smith, 2010) which is based on a combination of syntactic transformation rules implemented with *tregex* (Levy and Andrew, 2006) and statistical methods. To generate the *Compl(ex) Q(uestion)* in the *ComplQ;Expl* Dialogue Act (DA) sequence, we use a combination of the CMU tool and lexical transformation rules.³ The GEN example in Table 2 illustrates this: The input monologue has a Manner-Means relations between the nucleus ‘In September, Ashland settled the long-simmering dispute’ and the satellite ‘by agreeing to pay Iran 325 million USD’. The satellite is copied without alteration to the Explain dialogue act. The nucleus is processed by applying the following template-based rule:

Decl \Rightarrow How Yes/No Question(Decl)

In words, the input consisting of a declarative sentence is mapped to a sequence consisting of the word ‘How’ followed by a Yes/No-question (in this case “Did Ashland settle the long-simmering dispute in December?”) that is obtained with the CMU QG tool from the declarative input sentence. A similar approach is applied for the other relations (Attribution, Condition and Explanation-Reason) that can lead to a *ComplQ; Expl* dialogue act sequence (see Table 1).

Generally, sequences requiring only copying or canned text are labelled *d(irect)* in Table 1, whereas those requiring syntactic transformation are labelled *c(omplex)*.

³In contrast, the *ComplQ* in the DA sequence *Expl;ComplQ;Expl* is generated using canned text such as ‘Why?’ or ‘Why is that?’.

4 Evaluation

We evaluate the output generated with both complex and direct rules for the relations of Table 1.

4.1 Materials, Judges and Procedure

The input monologues were text excerpts from the Wall Street Journal as annotated in the RST Discourse Treebank⁴. They consisted of a single sentence with one internal relation, or two sentences (with no internal relations) connected by a single relation. To factor out the quality of the discourse annotations, we used the gold standard annotations of the Discourse Treebank and checked these for correctness, discarding a small number of incorrect annotations.⁵ We included text fragments with a variety of clause length, ordering of nucleus and satellite, and syntactic structure of clauses. Table 2 shows examples of monologue/dialogue pairs: one with a generated dialogue and the other from the corpus.

Our study involved a panel of four judges, each fluent speakers of English (three native) and experts in Natural Language Generation. We collected judgements on 53 pairs of monologue and corresponding dialogue. 19 pairs were judged by all four judges to obtain inter-annotator agreement statistics, the remainder was parcelled out. 38 pairs consisted of WSJ monologue and generated dialogue, henceforth GEN, and 15 pairs of CODA corpus monologue and human-authored dialogue, henceforth CORPUS (instances of generated and corpus dialogue were randomly interleaved) – see Table 2 for examples.

The two standard evaluation measures for language generation, accuracy and fluency (Mellish and Dale, 1998), were used: a) *accuracy*: whether a dialogue (from GEN or CORPUS) preserves the information of the corresponding monologue (judgement: ‘Yes’ or ‘No’) and b) *monologue and dialogue fluency*: how well written a piece of monologue or dialogue from GEN or CORPUS is. Fluency judgements were on a scale from 1 ‘incomprehensible’ to 5 ‘Comprehensible, grammatically correct and naturally sounding’.

⁴www.isi.edu/~marcu/discourse/Corpora.html

⁵For instance, in our view ‘without wondering’ is incorrectly connected with the attribution relation to ‘whether she is moving as gracefully as the scenery.’

GEN **Monologue**
 In September, Ashland settled the long-simmering dispute by agreeing to pay Iran 325 million USD.

Dialogue (*ComplQ; Expl*)

- A: How did Ashland settle the long-simmering dispute in December?
 B: By agreeing to pay Iran 325 million USD.

CORPUS **Monologue**
 If you say "I believe the world is round", the "I" is the mind.

Dialogue (*FactQ; FactA*)

- A: If you say "I believe the world is round", who is the "I" that is speaking?
 B: The mind.

Table 2: Monologue-Dialogue Instances

4.2 Results

Accuracy Three of the four judges marked 90% of monologue-dialogue pairs as presenting the same information (with pairwise κ of .64, .45 and .31). One judge interpreted the question differently and marked only 39% of pairs as containing the same information. We treated this as an outlier, and excluded the accuracy data of this judge. For the instances marked by more than one judge, we took the majority vote. We found that 12 out of 13 instances (or 92%) of dialogue and monologue pairs from the CORPUS benchmark sample were judged to contain the same information. For the GEN monologue-dialogue pairs, 28 out of 31 (90%) were judged to contain the same information.

Fluency Although absolute agreement between judges was low,⁶ pairwise agreement in terms of Spearman rank correlation (ρ) is reasonable (average: .69, best: .91, worst: .56). For the subset of instances with multiple annotations, we used the data from the judge with the highest average pair-wise agreement ($\rho = .86$)

The fluency ratings are summarised in Figure 1. Judges ranked both monologues and dialogues for

⁶For the four judges, we had an average pairwise κ of .34 with the maximum and minimum values of .52 and .23, respectively.

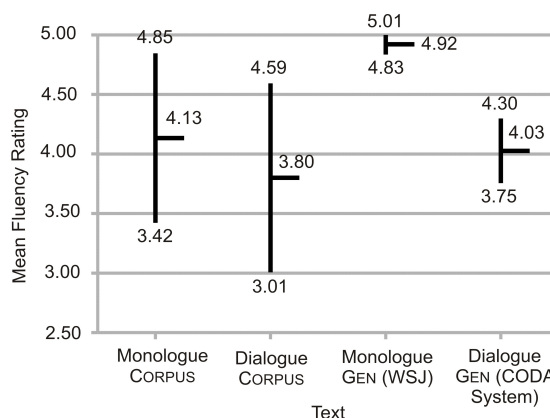


Figure 1: Mean Fluency Rating for Monologues and Dialogues (for 15 CORPUS and 38 GEN instances) with 95% confidence intervals

the GEN sample higher than for the CORPUS sample (possibly as a result of slightly greater length of the CORPUS fragments and some use of archaic language). However, the drop in fluency, see Figure 2, from monologue to dialogue is greater for GEN sample (average: .89 points on the rating scale) than the CORPUS sample (average: .33) (T-test $p < .05$), suggesting that there is scope for improving the generation algorithm.

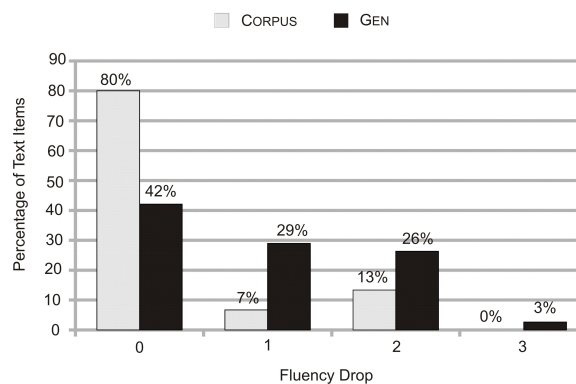


Figure 2: Fluency drop from monologue to corresponding dialogue (for 15 CORPUS and 38 GEN instances). On the x-axis the fluency drop is marked, starting from no fluency drop (0) to a fluency drop of 3 (i.e., the dialogue is rated 3 points less than the monologue on the rating scale).

Direct versus Complex rules We examined the difference in fluency drop between direct and complex rules. Figure 3 shows that the drop in fluency for dialogues generated with complex rules is higher than for the dialogues generated using direct rules (T-test $p < .05$). This suggests that use of direct rules is more likely to result in high quality dialogue. This is encouraging, given that Stoyanchev and Piwek (2010a) report higher frequencies in professionally authored dialogues of dialogue acts (*YNQ*, *Expl*) that can be dealt with using direct verbalisation (in contrast with low frequency of, e.g., *FactQ*).

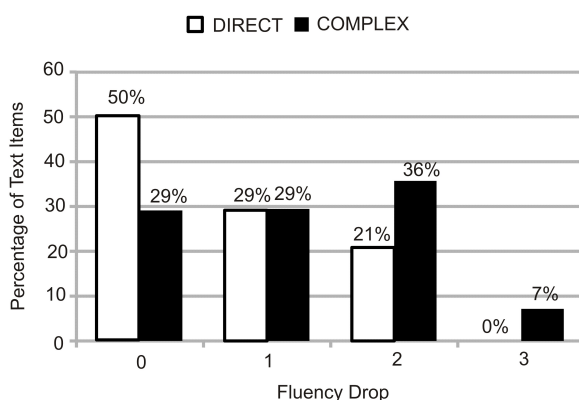


Figure 3: Decrease in Fluency Score from Monologue to Dialogue comparing Direct (24 samples) and Complex (14 samples) dialogue generation rules

5 Conclusions and Further Work

With information presentation in dialogue form being particularly suited for education and persuasion, the presented system is a step towards making information from text automatically available as dialogue. The system relies on discourse-to-dialogue structure rules that were automatically extracted from a parallel monologue/dialogue corpus. An evaluation against a benchmark sample from the human-written corpus shows that both accuracy and fluency of generated dialogues are not worse than that of human-written dialogues. However, drop in fluency between input monologue and output dialogue is slightly worse for generated dialogues than for the benchmark sample. We also established a difference in quality of output generated with complex versus direct discourse-to-dialogue rules, which can

be exploited to improve overall output quality.

In future research, we aim to evaluate the accuracy and fluency of longer stretches of generated dialogue. Additionally, we are currently carrying out a task-related evaluation of monologue versus dialogue to determine the utility of each.

Acknowledgements

We would like to thank the three anonymous reviewers for their helpful comments and suggestions. We are also grateful to our colleagues in the Open University's Natural Language Generation group for stimulating discussions and feedback. The research reported in this paper was carried out as part of the CODA research project (<http://computing.open.ac.uk/coda/>) which was funded by the UK's Engineering and Physical Sciences Research Council under Grant EP/G020981/1.

References

- E. André, T. Rist, S. van Mulken, M. Klesen, and S. Baldes. 2000. The automated design of believable dialogues for animated presentation teams. In Justine Cassell, Joseph Sullivan, Scott Prevost, and Elizabeth Churchill, editors, *Embodied Conversational Agents*, pages 220–255. MIT Press, Cambridge, Massachusetts.
- H. Bunt. 2000. Dialogue pragmatics and context specification. In H. Bunt and W. Black, editors, *Abduction, Belief and Context in Dialogue: Studies in Computational Pragmatics*, volume 1 of *Natural Language Processing*, pages 81–150. John Benjamins.
- J. Carletta, A. Isard, S. Isard, J. Kowtko, G. Doherty-Sneddon, and A. Anderson. 1997. The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23:13–31.
- L. Carlson and D. Marcu. 2001. Discourse tagging reference manual. Technical Report ISI-TR-545, ISI, September.
- M. Cavazza and F. Charles. 2005. Dialogue Generation in Character-based Interactive Storytelling. In *Proceedings of the AAAI First Annual Artificial Intelligence and Interactive Digital Entertainment Conference*, Marina Del Rey, California, USA.
- M. Core and J. Allen. 1997. Coding Dialogs with the DAMSL Annotation Scheme. In *Working Notes: AAAI Fall Symposium on Communicative Action in Humans and Machine*.

- S. Craig, B. Gholson, M. Ventura, A. Graesser, and the Tutoring Research Group. 2000. Overhearing dialogues and monologues in virtual tutoring sessions. *International Journal of Artificial Intelligence in Education*, 11:242–253.
- D. duVerle and H. Prendinger. 2009. A novel discourse parser based on support vector machines. In *Proc 47th Annual Meeting of the Association for Computational Linguistics and the 4th Int'l Joint Conf on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-IJCNLP'09)*, pages 665–673, Singapore, August.
- M. Heilman and N. A. Smith. 2010. Good question! statistical ranking for question generation. In *Proc. of NAACL/HLT*, Los Angeles.
- Huong T. Le and Geehta Abeyasinghe. 2003. A study to improve the efficiency of a discourse parsing system. In *Proceedings 4th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-03)*, Springer LNCS 2588, pages 101–114.
- J. Lee, F. Dinneen, and J. McKendree. 1998. Supporting student discussions: it isn't just talk. *Education and Information Technologies*, 3:217–229.
- R. Levy and G. Andrew. 2006. Tregex and tsurgeon: tools for querying and manipulating tree data structures. In *5th International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy.
- William C. Mann and Sandra A. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.
- M. Mateas and A. Stern. 2005. Structuring content in the faade interactive drama architecture. In *Proc. of Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, Marina del Rey, Los Angeles, June.
- C. Mellish and R. Dale. 1998. Evaluation in the context of natural language generation. *Computer Speech and Language*, 12:349–373.
- P. Piwek and S. Stoyanchev. 2010. Generating Expository Dialogue from Monologue: Motivation, Corpus and Preliminary Rules. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 333–336, Los Angeles, California, June.
- P. Piwek, H. Hernault, H. Prendinger, and M. Ishizuka. 2007. T2D: Generating Dialogues between Virtual Agents Automatically from Text. In *Intelligent Virtual Agents: Proceedings of IVA07*, LNAI 4722, pages 161–174. Springer Verlag.
- H. Prendinger and M. Ishizuka, editors. 2004. *Life-Like Characters: Tools, Affective Functions, and Applications*. Cognitive Technologies Series. Springer, Berlin.
- V. Rus, A. Graesser, A. Stent, M. Walker, and M. White. 2007. Text-to-Text Generation. In R. Dale and M. White, editors, *Shared Tasks and Comparative Evaluation in Natural Language Generation: Workshop Report*, Arlington, Virginia.
- S. Stoyanchev and P. Piwek. 2010a. Constructing the CODA corpus. In *Procs of LREC 2010*, Malta, May.
- S. Stoyanchev and P. Piwek. 2010b. Harvesting re-usable high-level rules for expository dialogue generation. In *6th International Natural Language Generation Conference (INLG 2010)*, Dublin, Ireland, 7-8, July.
- S. V. Suzuki and S. Yamada. 2004. Persuasion through overheard communication by life-like agents. In *Procs of the 2004 IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, Beijing, September.
- K. van Deemter, B. Krenn, P. Piwek, M. Klesen, M. Schroeder, and S. Baumann. 2008. Fully Generated Scripted Dialogue for Embodied Agents. *Artificial Intelligence Journal*, 172(10):1219–1244.