

# Improving Fluency in Narrative Text Generation With Grammatical Transformations and Probabilistic Parsing

Emily Ahn

Wellesley College

eahn@wellesley.edu

Fabrizio Morbini and Andrew S. Gordon

Institute for Creative Technologies, University of Southern California

morbini@ict.usc.edu, gordon@ict.usc.edu

## Abstract

In research on automatic generation of narrative text, story events are often formally represented as a causal graph. When serializing and realizing this causal graph as natural language text, simple approaches produce cumbersome sentences with repetitive syntactic structure, e.g. long chains of “because” clauses. In our research, we show that the fluency of narrative text generated from causal graphs can be improved by applying rule-based grammatical transformations to generate many sentence variations with equivalent semantics, then selecting the variation that has the highest probability using a probabilistic syntactic parser. We evaluate our approach by generating narrative text from causal graphs that encode 100 brief stories involving the same three characters, based on a classic film of experimental social psychology. Crowdsourced workers judged the writing quality of texts generated with ranked transformations as significantly higher than those without, and not significantly lower than human-authored narratives of the same situations.

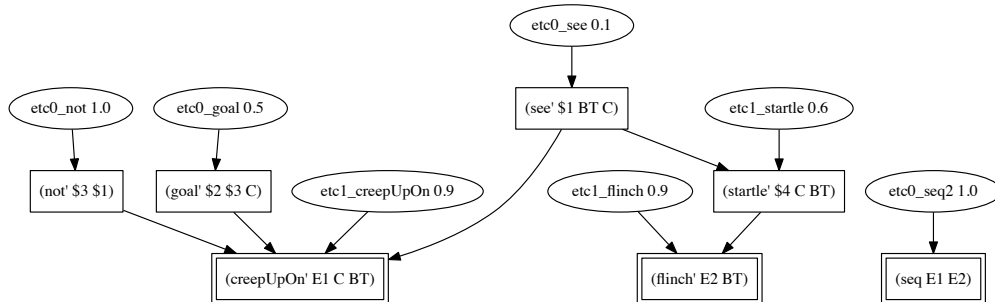
## 1 Narrative Text Generation

Across several academic disciplines, it has become common to represent narratives as causal graphs. In the *causal network model* of psychologists Trabasso and van den Broek (1985), vertices in the graph structure represent settings, events, goals, attempts and outcomes of a narrative, linked via directed edges that encode cause/effect relationships. In computer science, similar causal graphs have been

used to model and manipulate narrative elements including suspense (Cheong and Young, 2014), conflict (Ware and Young, 2011), flashback and foreshadowing (Bae and Young, 2008). Elson (2012) elaborates the causal network model by relating it to both the temporal ordering of story-world events and an author’s textual realization, creating a three-layer Story Intention Graph.

Causal graph representations of narrative create new opportunities for natural language generation (NLG) of narrative text. For example, Lukin et al. (2015) describe a narrative NLG pipeline for Story Intention Graphs, generating variations of an original text that can be parameterized for particular discourse goals. When serializing and realizing a causal graph structure as natural language text, some care must be taken to avoid the generation of cumbersome sentences with repetitive syntactic structure, e.g. as a long chain of “because” clauses. Lukin et al. (2015) directly compared readers’ overall preferences for certain causal connectives over others, finding that no single class of variations will produce sentences that are preferable to a human author’s stylistic choices.

We hypothesize that the policies used by native speakers to select among lexical-syntactic variations are complex and content-dependent, and are best described in statistical models trained on natural language corpora. In this paper, we explore a new approach to narrative NLG that integrates rule-based and statistical methods to produce fluent realizations of storylines encoded as causal graphs. Beginning with the output of a simple baseline system, we show that the fluency of narrative text generated



**Figure 1:** Causal graph output for Triangle-COPA question 1, where a circle (C) creeps up on a big triangle (BT), who then flinches.

from causal graphs can be improved by applying rule-based grammatical transformations to generate many sentence variations with equivalent semantics, then selecting the variation that has the highest probability using a probabilistic syntactic parser. Our software implementation is available online.<sup>1</sup>

## 2 Triangle-COPA Causal Graphs

As a corpus of causal graphs for use as input for our NLG experiments, we used a set of solutions to the 100 interpretation problems in the Triangle-COPA evaluation.<sup>2</sup> This evaluation set is based on a film from 1958 created by Fritz Heider and Marianne Simmel for use in early social psychology experiments (Heider and Simmel, 1944), depicting the movements of geometric shapes (two triangles and a circle) moving in and around a box with a door. The Triangle-COPA evaluation consists of 100 short movies in the same style, with both natural language and formal descriptions of their content. The evaluation tests a system’s ability to select the most human-like interpretation of the observed behavior among two choices (Maslan et al., 2015).

Gordon (2016) demonstrated that Triangle-COPA questions could be solved by automatically constructing causal graphs that explain the behavior of the moving shapes in terms of their underlying goals, emotions, and social relationships. The approach pairs a hand-authored knowledge base of commonsense causal axioms with an abductive reasoning engine that finds the most-probable set of assumptions that logically entail the observed behavior.<sup>3</sup> Forward-chaining from these assumptions us-

ing the knowledge base axioms produces a directed causal graph, where the goals, emotions, and social relationships of the characters are intermediate inferences in proving the observations.

In our own research, we generated causal graphs for each of the 100 Triangle-COPA questions using the abduction engine and knowledge base of Gordon (2016), for use as input in our NLG experiments. An example causal graph solution appears in Figure 1 (for Triangle-COPA question 1). In our evaluations, we also used the human-authored narrative included with each question in the Triangle-COPA question set, which represents the realization of the writer’s own interpretation of the events depicted in a given Triangle-COPA movie. For the same movie that produced the interpretation represented in Figure 1, the following is the human-authored narrative:

*A circle stealthily stalks a big triangle. It does not want to be seen so it moves very slowly and quietly and then suddenly startles the triangle.*

## 3 Baseline NLG System

We developed a baseline NLG system that transforms our causal graphs into narrative texts. Our approach was to first divide the input graph into sections containing exactly one timeline event (*creep-up-on* and *flinch*, in Figure 1) along with its causal antecedents. Each of these sections becomes a sentence in the generated output, ordered by any sequence information provided in the input (*seq* in Figure 1). Each sentence is structured as a chain of “because” clauses, beginning with the timeline event and followed by each of its causal antecedents. These structures are then realized as text using the

<sup>1</sup><https://github.com/fmorbini/hsit-generation>

<sup>2</sup><https://github.com/asgordon/TriangleCOPA>

<sup>3</sup><https://github.com/asgordon/EtcAbductionPy>

SimpleNLG engine (Gatt and Reiter, 2009) with the support of a custom lexicon for the specific predicates used in Triangle-COPA’s representations.

Below is an example of the output of this Baseline NLG system, generated from the causal graph depicted in Figure 1. As expected, this text exhibits cumbersome phrasing and repetitive structure.

*The circle creeps up on the big triangle because the circle wants that the big triangle does not see the circle. The big triangle flinches because the circle startles the big triangle because the big triangle sees the circle.*

#### 4 Grammatical Transformations

We sought to improve the fluency of our baseline NLG system by generating many variations of each sentence through domain-independent grammatical transformations, then ordering these variations to select the best one. In this section, we describe the set of 24 hand-authored rules for grammatical transformations used in our experiments, of 7 types:

**Sentential arguments:** These transformations improve the fluency of verb phrases with sentential arguments. Example input: *A wants that B does C.* Output: *A wants B to do C.*

**Causality:** These transformations realize the causality relation in ways other than the default “because” connective. Example input: *A does B because A wants C.* (where the subject of C is A) Output: *A does B to C.*

**Conjunction introduction:** These transformations simplify neighboring structures that share some components (e.g. a subject). Example input: *A does B and A does C.* Output: *A does B and C.*

**Repetitions:** Identical timeline events in sequences are combined. Example input: *A does B. A does B. A does B.* Output: *A does B repeatedly.*

**Intermediate deletion:** These transformations remove intermediate vertices in causal chains, under the assumption that some causal links are intuitive and can be left implicit. Example input: *A ignores B because A dislikes B because B annoys A.* Output: *A ignores B because B annoys A.*

**Pronoun introduction:** These transformations replace proper nouns with pronouns when it is unambiguous to do so. Example input: *A ignores B*

*because B annoys A.* Output: *A ignores B because A annoys it.*

**Lexical fixes:** These transformations handle special cases of lexical-syntactic patterns not easily handled by the realization engine. Example input: *A knocks the door in order to B.* Output: *A knocks on the door in order to B.*

Each of the 24 hand-authored rules in our system is applied recursively and exhaustively to the syntactic structures used in our baseline NLG system, generating tens to hundreds of variations for each input sentence. We explored the use of a large-scale paraphrase database (Ganitkevitch et al., 2013) as a source of transformation rules, but found that it contained none that were equivalent to those in our hand-authored set. The advantage our hand-authored rules is that they strictly preserve the semantics of the original input.

#### 5 Probabilistic Parsing

To select the best variation for each sentence in the output narrative, we parse each variant using a probabilistic syntactic parser and rank according to the probability of the generated parse tree. For this purpose, we use the constituency parser of Charniak and Johnson (2005) without the built-in reranker, using model `SANCL2012-Uniform`. Each variant is grammatically correct, so our interest is solely the assigned probability score for the typicality of the lexical-syntactic structure in the training data. Here the parser serves the same role as n-gram language models in machine translation or speech recognition systems, but should be better suited for our task where intra-sentence long-range dependencies are factors in the quality of the text. We investigated whether normalizing these scores by sentence length would improve rankings, but our evaluations here are based on unnormalized probability scores.

After selecting the most-probable variant for each sentence, we assembled a final narrative for each of the 100 causal graphs. For example, the graph in Figure 1 produces the following output:

*The circle creeps up on the big triangle because she does not want him to see her. He flinches because he sees her.*

## 6 Evaluation

We evaluated the quality of our narrative NLG approach by soliciting ratings of writing quality from crowdsourced annotators, comparing the output of our system, our baseline NLG system, and original human-authored narratives for each of the 100 questions in the Triangle-COPA question set. In each annotation task, the annotator watched the short movie associated with a given question, read the text associated with the question randomly selected from our three conditions, then rated the writing quality of the text on a 5-point Likert scale - from (1) *Horrible gibberish* to (5) *Excellent, professional quality*. In addition, we asked raters to answer a factual multiple-choice question about each movie to validate their effort on this crowdsourced task. After filtering annotators who failed this validation task, we analyzed 717 ratings evenly distributed across the three conditions and 100 questions, shown in Table 1. Significant gains in quality ratings were observed for our approach over the Baseline NLG system. The differences observed between human-authored narratives and our system were not significant.

Condition	Ratings	Mean score (1-5)
Human authored	233	3.69
Our system	236	3.59 *
Baseline NLG	248	3.11

Table 1: Ratings of writing quality. (\*) significant at  $p < 0.05$

## 7 Conclusions

This research demonstrates that high-quality textual narratives can be generated from causal graph representations of stories. The use of hand-authored grammatical transformation rules helps ensure that all textual variations retain the semantics of the original input, while probabilistic parsing helps identify the variation that corresponds best to the structures produced by native speakers.

In our study, the input causal graphs were also automatically generated, identified as the most probable explanations of series of observable events using logical abduction. Having combined automated *interpretation* with automated *narrative generation*, we now wonder if automated *perception* algorithms could serve as the input to similar pipelines

to enable future systems to generate human-like narratives of the events in real-world situations.

## Acknowledgments

This research was supported by the Office of Naval Research, grant N00014-13-1-0286. This material is based upon work supported by the National Science Foundation under Grant No. 1263386.

## References

- Byung-Chul Bae and R Michael Young. 2008. A use of flashback and foreshadowing for surprise arousal in narrative using a plan-based approach. In *International Conference on Interactive Digital Storytelling*.
- Eugene Charniak and Mark Johnson. 2005. Coarse-to-fine n-best parsing and maxent discriminative reranking. In *43rd Annual Meeting on Association for Computational Linguistics*.
- Yun-Gyung Cheong and R Michael Young. 2014. Suspenser: A story generation system for suspense. *Transactions on Computational Intelligence and Artificial Intelligence in Games*.
- David Elson. 2012. *Modeling Narrative Discourse*. Ph.D. thesis, Columbia University.
- Juri Ganitkevitch, Benjamin Van Durme, and Chris Callison-Burch. 2013. PPDB: The paraphrase database. In *2013 Meeting of the North American Association for Computational Linguistics*.
- Albert Gatt and Ehud Reiter. 2009. SimpleNLG: A realisation engine for practical applications. In *12th European Workshop on Natural Language Generation*.
- Andrew S. Gordon. 2016. Commonsense interpretation of triangle behavior. In *30th AAAI Conference on Artificial Intelligence*.
- Fritz Heider and Marianne Simmel. 1944. An experimental study of apparent behavior. *The American Journal of Psychology*, 57(2):243–259.
- Stephanie Lukin and Marilyn Walker. 2015. Generating sentence planning variations for story telling. In *16th Annual SIGdial meeting on Discourse and Dialogue*.
- Nicole Maslan, Melissa Roemmele, and Andrew S. Gordon. 2015. One hundred challenge problems for logical formalizations of commonsense psychology. In *12th International Symposium on Logical Formalizations of Commonsense Reasoning*.
- Tom Trabasso and Paul Van Den Broek. 1985. Causal thinking and the representation of narrative events. *Journal of memory and language*, 24(5):612–630.
- Stephen Ware and R Michael Young. 2011. Cprocl: A narrative planner supporting conflict. In *7th International Conference on Artificial Intelligence and Interactive Digital Entertainment*.