

## **UC Merced**

### **Proceedings of the Annual Meeting of the Cognitive Science Society**

#### **Title**

Can computers tell a story? Discourse Structure in Computer-generated Text and Humans

#### **Permalink**

<https://escholarship.org/uc/item/2276k59j>

#### **Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

#### **ISSN**

1069-7977

#### **Authors**

Wey, Alex  
Hemmatian, Babak  
Avram, Rachel  
et al.

#### **Publication Date**

2021

Peer reviewed

# Can computers tell a story? Discourse Structure in Computer-generated Text and Humans

**Alex Wey**

Brown University, Providence, Rhode Island, United States

**Babak Hemmatian**

Brown University, Providence, Rhode Island, United States

**Rachel Avram**

Brown University, Providence, Rhode Island, United States

**Sheridan Feucht**

Brown University, Providence, Rhode Island, United States

**Kate Spitalnic**

University of Sussex, Falmer, Brighton, United Kingdom

**Muskaan Garg**

Brown University, Providence, Rhode Island, United States

**Carsten Eickhoff**

Brown University, Providence, Rhode Island, United States

**Ellie Pavlick**

Brown University, Providence, Rhode Island, United States

**Björn Sandstede**

Brown University, Providence, Rhode Island, United States

**Steven Sloman**

Brown University, Providence, Rhode Island, United States

## Abstract

Text-generation algorithms like GPT-2 (Radford et al., 2019) and GPT-3 (Brown et al., 2020) produce documents which resemble coherent human writing. But no study has compared the discourse linguistics features of the artificial text with that of comparable human content. We used a sample of Reddit and news discourse as prompts to generate artificial text using fine-tuned GPT-2 (Grover; Zellers et al., 2019). Blind annotators identified clause-level discourse features (e.g., states and events; Smith, 2003), and coherence relations (e.g., contrast; Wolf and Gibson, 2005) in prompts and generated text. Comparing the >20000 clauses, Grover recreates human word co-occurrence patterns and clause types across discourse modes. However, its coherence relations have shorter length and lower quality, with many nonsensical instances. Therefore, annotators could perfectly guess the human/algorithmic source of documents. Using a corresponding GPT-3 sample, we discuss aspects of generation that have and have not improved since Grover.