# Automated Assistance for Creative Writing with an RNN Language Model

**Melissa Roemmele and Andrew S. Gordon**
Institute for Creative Technologies, University of Southern California, Los Angeles, CA
roemmele@ict.usc.edu, gordon@ict.usc.edu

## ABSTRACT
This work demonstrates an interface, Creative Help, that assists people with creative writing by automatically suggesting new sentences in a story. Authors can freely edit the generated suggestions, and the application tracks their modifications. We make use of a Recurrent Neural Network language model to generate suggestions in a simple probabilistic way. Motivated by the theorized role of unpredictability in creativity, we vary the degree of randomness in the probability distribution used to generate the sentences, and find that authors' interactions with the suggestions are influenced by this randomness.

## ACM Classification Keywords
H.5.2 Information Interfaces and Presentation (e.g. HCI): User Interfaces; I.2.7 Natural Language Processing

## Author Keywords
Story Generation; Computational Creativity; Writing Support

## INTRODUCTION
At the intersection between natural language generation, computational creativity, and human-computer interaction research is the vision of automated tools that collaborate with people in authoring creative text. The recent application Creative Help [4] explores this vision for story writing. The interface is simple: authors type *\help* to generate a suggestion for a new sentence in an ongoing story, which they can then edit. The application tracks authors' edits to suggestions as a strategy for evaluation. Figure 1 shows an example with the suggestion returned by the help request underlined. In [4], a nearest-neighbors similarity approach was used to produce Creative Help suggestions by retrieving sentences from a large story corpus. The current demo applies a different generation approach that dynamically generates novel sentences word-by-word. Some researchers have theorized that randomness plays a large role in human creativity, on the basis that creativity involves making sense out of unpredictable combinations of ideas. Accordingly, given the difficulty of developing systems that inherently model human creativity, it may be desirable to leverage randomness as a way of simulating creativity [1]. In the context of creative writing assistance, unpredictable word combinations may serendipitously present a novel idea to the author. Of course, completely random sentences will most often be unintelligible, so they must be constrained to some degree. Based on this, we explore a Recurrent Neural Network language model (RNN LM) [3] for this task, which learns a conditional probability distribution of each word occurring in a story given the words that precede it. See [5] for details about the specific architecture of this model. The RNN LM was trained on 8032 fiction books in the BookCorpus[1]. To generate a new sentence in response to a help request, the existing story is provided as input to the model, and the model returns a probability distribution for all potential first words in the next sentence. We randomly sample a word from this distribution, append it to the new sentence, and iteratively continue doing this until an end-of-sentence punctuation mark is generated. In the experiment described below, we manipulated the degree of randomness in the generated sentences by using a 'temperature' variable that adjusts the probability distribution. [2] used this same approach to vary generation for a task similar to ours, but did not conduct a user study to assess its effect. Lowering the temperature value skews the distribution so that the probability is concentrated more densely under fewer words. Consequently, it causes the model to generate less random, more predictable sequences according to those observed in the training data.


**Figure 1. Creative Help interface with generated sentence**

## EXPERIMENT AND RESULTS
We set up a study where we recruited people via social media, email, and Amazon Mechanical Turk to interact with Creative Help. Participants were instructed to write a story about any topic. They were told the objective of the task was to experiment with asking for help but that they were not required to make a certain number of requests. They could choose to edit, add to, or delete a suggestion just like any other text in their story. The site randomly assigned the user to one of

[1] yknzhu.wixsite.com/mbweb



*IUI'18 Companion* March 7–11, 2018, Tokyo, Japan



ACM ISBN 978-1-4503-5571-1/18/03.
DOI: https://doi.org/10.1145/3180308.3180329

| | Temperature | |
|---|---|---|
| | 0.6 | 1.0 |
| 1. On average, how grammatically correct were the suggested sentences? | 3.42 | 2.35 |
| 2. On average, how coherent were the suggestions with the overall story? | 2.46 | 1.97 |
| 3. On average, how entertaining were the suggestions? | 3.46 | 3.56 |
| 4. On average, how original were the suggestions? | 3.31 | 3.65 |
| 5. Overall, did the suggestions make the story easier to write? | 3.14 | 2.29 |
| 6. Overall, how much did the suggestions influence your writing? | 3.59 | 2.60 |
| 7. Overall, how helpful were the suggestions for writing the story? | 3.20 | 2.43 |

**Table 1. Average ratings for the questionnaire-based metrics compared by temperature setting, with ratings reported on a 1-5 Likert scale**

two temperature settings, 0.6[2] or 1.0[3]. There was no theoretical guideline for comparing these particular values, but intuitively, they explore a trade-off between unpredictability and intelligibility. The 1.0 setting is more likely to generate novel sentences it has not observed during training, but at the risk of decreased intelligibility. By constraining some of the randomness in the probability distribution, the 0.6 temperature has a better chance of generating intelligible sentences, but they may be more predictable and thus less interesting in terms of creativity. Each time the user typed *help*\, a suggestion was generated according to the assigned temperature. After 15 minutes, the user was provided with a link to the questionnaire, but they could continue writing with no maximum time limit. Ultimately, 139 users participated in the task, with 70 assigned to the lower temperature (0.6) condition and 69 to the higher temperature condition (1.0). This resulted in suggestion-modification pairs for 940 help requests (514 for lower temperature and 426 for the higher temperature).

The questionnaire asked authors to provide ratings of their interaction with the application across different dimensions. Table 1 shows the questions with authors' average responses reported on a 1-5 Likert scale, compared by the temperature condition they observed. The gray rows indicate statistically significant differences between temperatures, determined through two-sample Monte Carlo permutation tests with $p < 0.025$. Relative to the higher temperature sentences, authors judged the lower temperature sentences as more grammatical and coherent. The lower temperature sentences also eased the writing of the story more, influenced its content more, and were more helpful overall. On entertainment and originality, the suggestions were comparable. It should be noted that even though the lower temperature suggestions were more favorable than the higher temperature ones, objectively their ratings were not particularly high for any dimension. In particular, the coherence of the suggestions for both temperatures was low.

We also evaluated authors' attitudes towards the suggestions in terms of how much they modified them, based on the idea that more helpful suggestions will be edited less. To do this, we computed the Levenshtein edit distance similarity between each suggestion and corresponding modification. The mean similarity score was 0.746 for the lower temperature suggestions versus 0.635 for the higher temperature ones, meaning that authors made significantly fewer changes to the lower temperature suggestions. Of course, it is possible for a suggestion to be heavily edited but also creatively stimulating. However, given that the questionnaire metrics mostly favor the lower temperature suggestions, it appears that overall authors indeed performed fewer edits to suggestions they found more helpful.

[2]https://fiction.ict.usc.edu/creativehelp/?gen_temp=0.6
[3]https://fiction.ict.usc.edu/creativehelp/?gen_temp=1.0

## CONCLUSION

In this work, we demonstrate an interface that provides automated support for story writing, where suggestions for new sentences in a story are generated by an RNN LM. We conducted an empirical evaluation of authors' interactions with the interface, where we specifically varied the level of unpredictability in the generated sentences and observed an effect on users' attitude and behavior towards the sentences. The role of randomness in simulating creativity is worth pursuing further. As models become more sophisticated, there is an important question of which components benefit most from being unpredictable. For example, authors may want the system to exhibit a fixed grammatical style while promoting unpredictability in generating plot events. More generally, additional work is needed to define precisely which aspects of writing should be supported by these interfaces.

## ACKNOWLEDGMENTS

The projects or efforts depicted were or are sponsored by the U.S. Army. The content or information presented does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

## REFERENCES

1. Margaret A Boden. 2004. *The creative mind: Myths and mechanisms*. Psychology Press.
2. Enrique Manjavacas, Folgert Karsdorp, Ben Burtenshaw, and Mike Kestemont. 2017. Synthetic Literature: Writing Science Fiction in a Co-Creative Process. In *CC-NLG 2017*. 29–37.
3. Tomáš Mikolov, Martin Karafiát, Lukáš Burget, Jan Černockỳ, and Sanjeev Khudanpur. 2010. Recurrent Neural Network based Language Model. In *INTERSPEECH 2010*. 1045–1048.
4. Melissa Roemmele and Andrew S Gordon. 2015. Creative Help: A Story Writing Assistant. In *ICIDS 2015*. Springer International Publishing.
5. Melissa Roemmele, Andrew S Gordon, and Reid Swanson. 2017. Evaluating Story Generation Systems Using Automated Linguistic Analyses. In *SIGKDD 2017 Workshop on Machine Learning for Creativity*.