

Mapping the Creativity Continuum: Automatic Measures of Creative Language Across Text Types

Creativity remains difficult to define and even harder to measure objectively. While certain text types, such as novels and user-generated reviews, are commonly perceived as creative, it is unclear whether this intuition can be supported by quantitative linguistic evidence. This study operationalises creativity as a quantifiable construct using automatic linguistic metrics and investigates whether these measures capture differences in creative language use across text types. It examines whether parameters of online reviews are closer to literary texts, with informative texts ranking as the least creative, proving that text types can be mapped along a hypothesised creativity continuum.

Previous research has demonstrated that creativity can be operationalised through linguistic features such as surprisal and semantic distance, which reflect novelty in language use (Beaty and Johnson, 2021, Bunescu and Uduehi, 2022). The present study extends this line of research by combining multiple computational metrics. Creativity is assessed using surprisal and semantic distance, alongside other potentially useful indicators including hapax legomena, n-gram coverage, and collocational strength measures. Judgements of creative potential are approximated through large language model ratings following a standardised Likert-scale prompting protocol (Guerberof-Arenas and Toral, 2022, Orwig et al., 2024).

The analysis is based on a pilot corpus of four English-language subcorpora: news articles, professional film reviews, user-generated book reviews and science fiction literary texts. Each subcorpus contains 20 texts and is annotated with metadata.

Preliminary results indicate that the examined text types can be positioned along a creativity continuum. News articles exhibit the lowest levels of creativity, followed by professional reviews, while user-generated reviews and literary texts occupy the higher end of the scale. These findings provide initial support for the claim that creative language can be measured using automatic metrics.

The study contributes to research in computational linguistics by further operationalising creativity as a measurable linguistic construct and establishing a methodological basis for analysing user-generated content as creative language. It also advances broader discussions on the validity of automated creativity assessment. The proposed framework provides a basis for investigating how creative text types are processed in translation, particularly in assessing the performance of machine translation and AI systems in handling creative language.

References:

- Beaty, R. E., & Johnson, D. R. 2021. Automating creativity assessment with SemDis: An open platform for computing semantic distance. *Behavior Research Methods*, 53(2), 757–780.
- Bunescu, R. C., & Uduehi, O. O. 2022. Distribution-based measures of surprise for creative language: Experiments with humor and metaphor. In D. Ghosh, B. Beigman Klebanov, S. Muresan, A. Feldman, S. Poria, & T. Chakrabarty (Eds.), *Proceedings of the 3rd Workshop on Figurative Language Processing (FLP)* (pp. 68–78). Association for Computational Linguistics.
- Guerberof-Arenas, Ana and Antonio Toral. 2022. “Creativity in translation: Machine translation as a constraint for literary texts”, *Translation Spaces* 11, 2: 184-212.
- Orwig, W., Edenbaum, E. R., Greene, J. D., & Schacter, D. L. 2024. The language of creativity: Evidence from humans and large language models. *The Journal of Creative Behavior*, 58(1), 128–136.

Wordcount: 353

Keywords: creativity, computational linguistics, semantic distance, surprisal, user-generated content