

# "What the <mask> did you call me?": How people recognize unrecognized words

Emily Zou Jennifer Pan

Communication Department, Stanford University

easyone@stanford.edu

## 1. The Puzzle

► People use unrecognized words online even without explicit definitions. How do people know what to do with undefined words?

- 1 Analyze conversations on online discussion forum, /pol/ from Jan 1- Mar 1 2023 (10,016,143 posts)
- 2 Classified words that appeared at least once a day as recognized or not in English language

## 2. Most recognized and unrecognized words (re)appear predictably, some in bursts

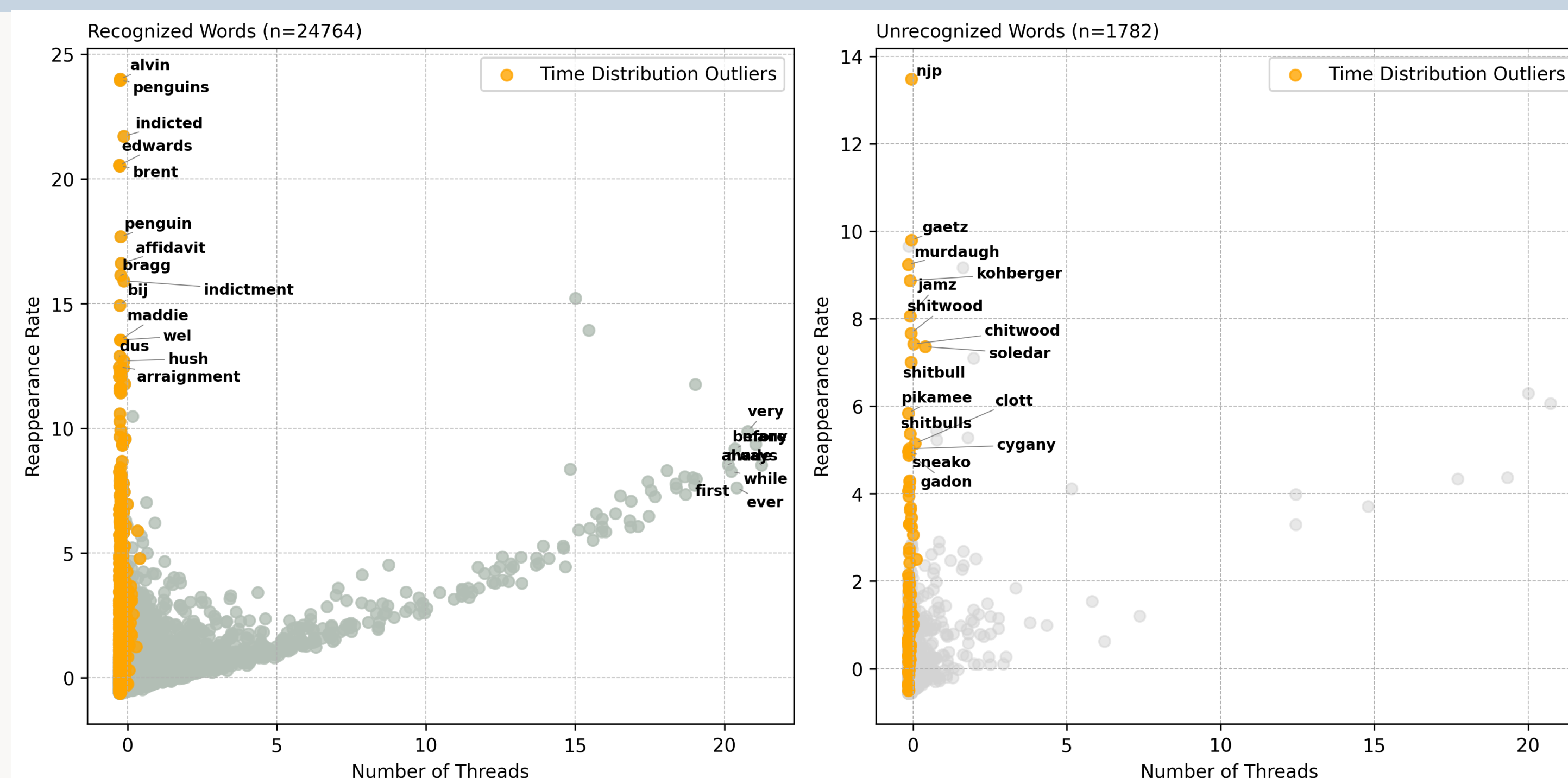


Figure: Comparison of word reappearance within (reappearance rate) vs. across conversation threads

- Most recognized & some unrecognized words appear normally, at similar rates within and across threads (normal words)
- Some recognized and unrecognized words appear in bursts, i.e., appears frequently (high reappearance rate) within one thread (burst words)
- Does this tell us anything about the *meaning* of the words?

## 3. Infer meaning from context of unrecognized words

- Use masked-word prediction test (RoBERTa) to guess words
- Label guessed words with part of speech (POS) tag
- Model performs well on words with known definitions (38% exact word match, 87% POS match)

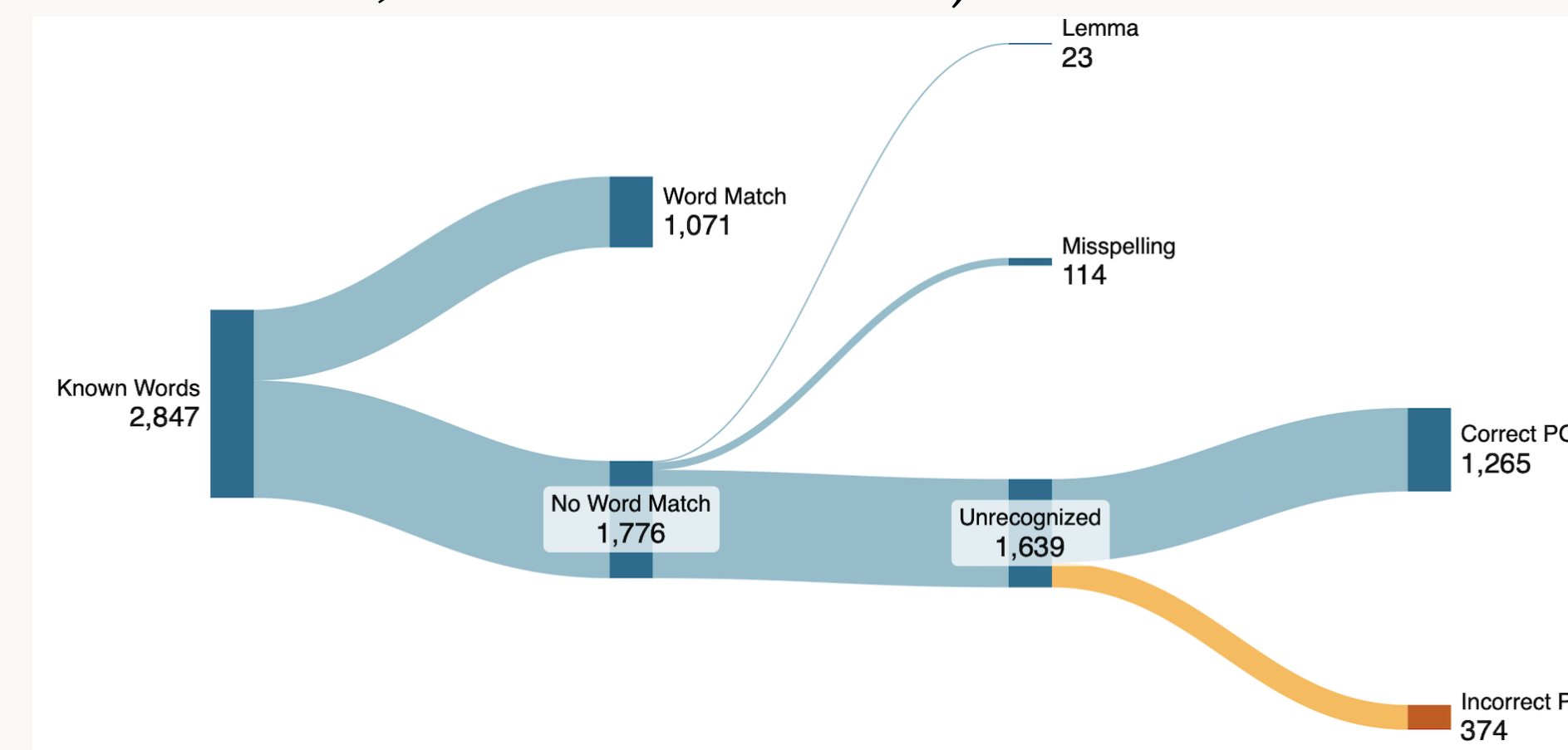


Figure: Model performance

## 4. Usage differentiates meaning for unrecognized words, but not recognized words

- Cosine similarity to measure 'cohesion' of its distribution of guessed words, which we take as semantic meaning.
- Recognized words have higher cosine similarity (more coherent meaning)
- Unrecognized burst words have greater coherence in meaning than unrecognized normal words

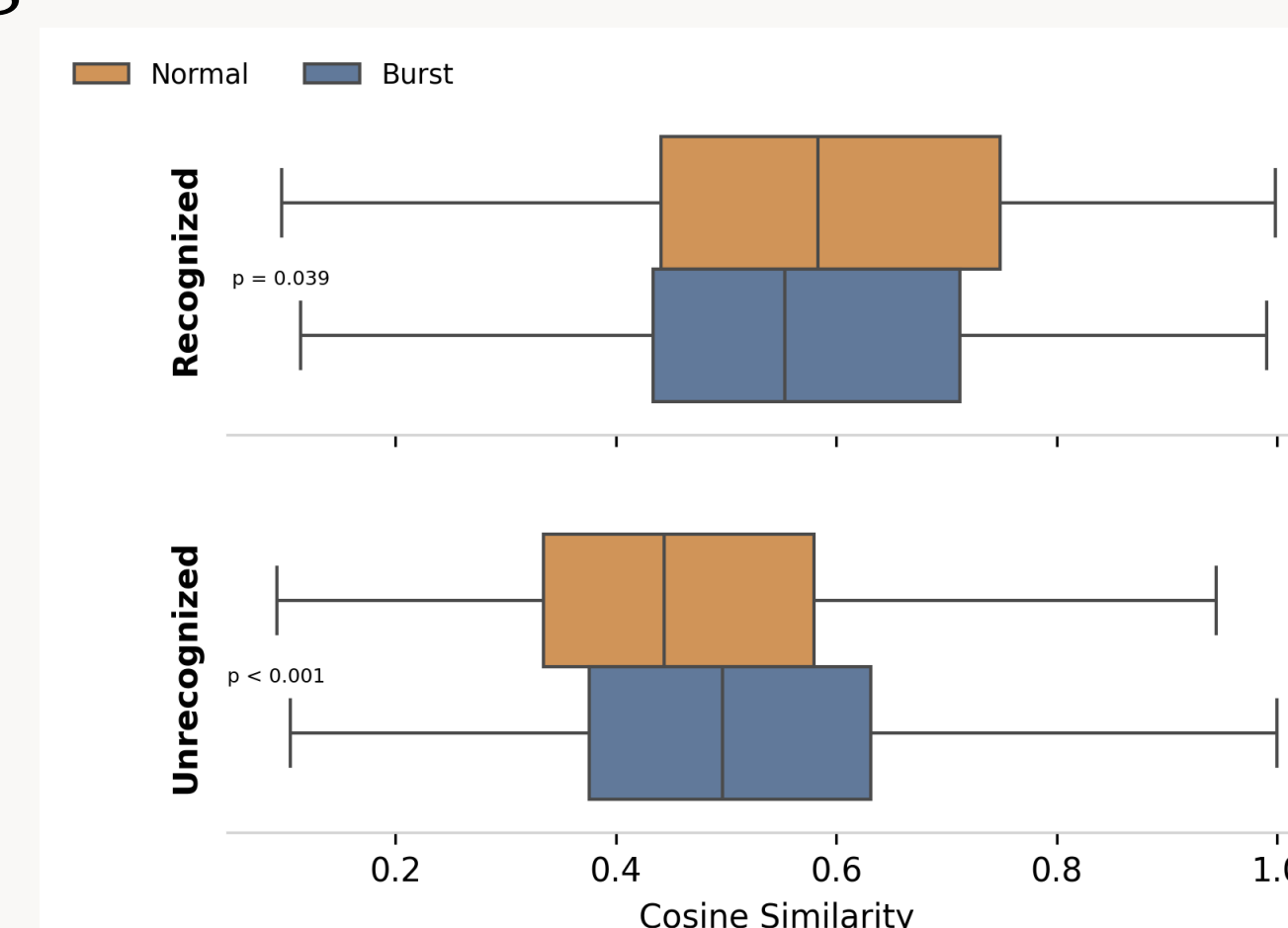


Figure: Usage differentiates meaning for unrecognized words

## 5. Unrecognized burst word guesses are more diverse across words

- More unique word guesses for unrecognized burst words (distribution of word guesses more closely centered around one) than unrecognized normal words
- Potentially implies guesses are more meaningful for burst words

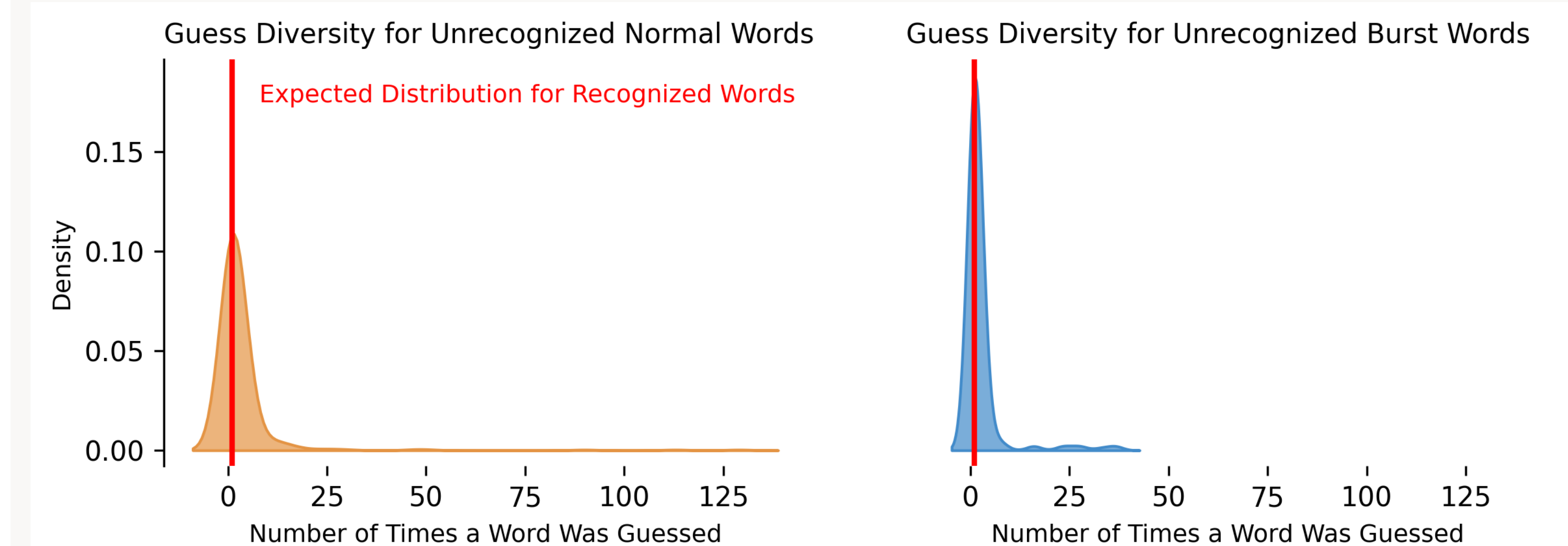


Figure: Burst words are more diverse

## 6. Discussion

- **Summary:** There are unrecognized normal and burst words, but burst words have more cohesive and unique meaning.
- **Substantive contribution:** people use unrecognized words online even without explicit definitions.
  - 1 Explicit meaning does not undergird communication
  - 2 (Online) social contexts create new, context specific language
- **Methodological contribution:** 'Discovering' vocabulary in NLP pipeline
- **Future work:**
  - 1 What do these words uniquely allow people to communicate?
  - 2 Why do people use these words?