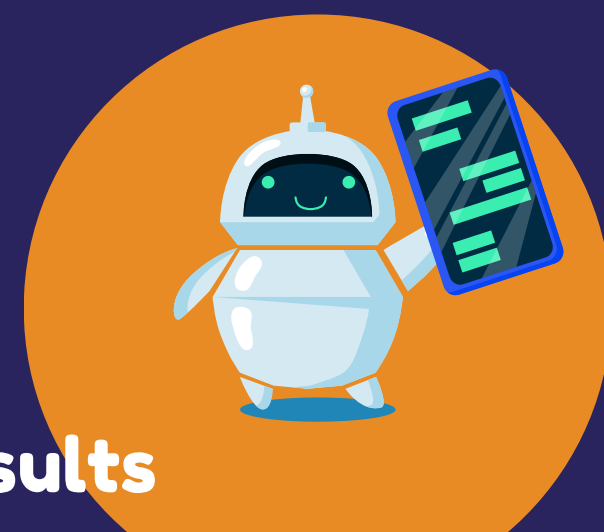# Prompting ChatGPT to Draw Morphological Connections for New Word Comprehension

Bianca Madalina Zgreaban
madalinazgreaban@gmail.com
Rishabh Suresh
r.suresh@students.uu.nl

Utrecht University

Background and RQ

- LLMs are more powerful, but retraining consumes energy [1]. Solution: prompt engineering [3], [9];
- Prompt engineering: better prompts > better performance [3], [9].
- **Case study**: ability to comprehend unfamiliar words using morphological connections in a LLM.
- The model: ChatGPT [7], fine-tuned on GPT-3.5.
    - Why? Task-agnostic! Limited studies for its neologism comprehension [5], [6].

- **H**ypothesis: *ChatGPT will perform better when the prompt provides more contextual information.*

Methods

- 24 prompts to test – 6 distinct *prompt patterns* x 4 conditions for context (e.g. see **1**):
    - *Prompt patterns* allowed us to be systematic and thorough – intent, motivation, structure, key ideas, consequences, and example from [8];
    - The prompts vary in detail, constraints, and domain-specificity [3];
    - Context – presence and absence of two task-specific keywords: 'morpheme' and 'new';
- 10 nonce test words of 1 existing word + 1 derivational suffix:

signatorily, assemblyless, benchish, delve able, lunchify, palatialise, violinous, musksome, containary, shallowen

    - 'Productive' suffixes (*-ly, -less, -ish, -able, -ify*) or 'unproductive' (*-ise, -ous, - some, -ary, and -en*) from [4].

- In total 240 trials conducted over two weeks, one chat session per trial

Define Pattern

1. Define...

2. Define the *new* word ...

3. Define the word ... considering its *morphemes*.

4. Define the *new* word ... considering its *morphemes*.

1

Results

Interaction between "new", "morpheme", and the productivity of the morpheme in the test word for humanlikeness

Interaction between "new", "morpheme", and the productivity of the morpheme in the test word for plausibility

BEST PROMPT?

No single prompt was statistically the "best"! But: the 'Word Generator Persona' prompt (with *new* and *morpheme)* had the most humanlike and plausible output

Scoring and Analysis

Two binary criteria: 1)*plausibility* and 2)*humanlikeness*
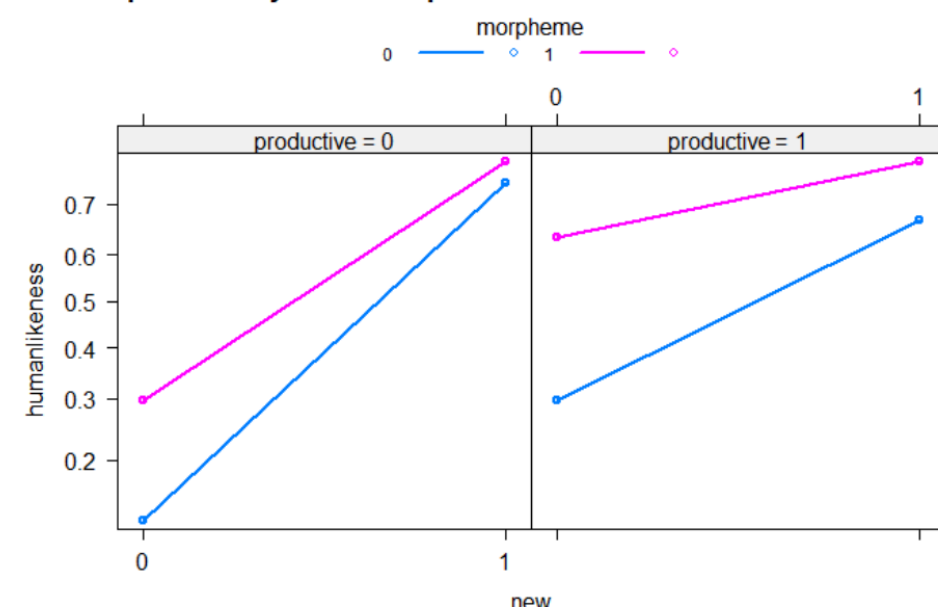
Is the definition:
- 1) based on the meaning of the root and the meaning of the suffix?
- 2) similar to the definitions of 11 human participants who also defined the test items? (e.g. See **2**)

- Analysis: generalised linear mixed-effect regression models for each criterion: morpheme*new*productive interaction, and by-prompt type intercepts
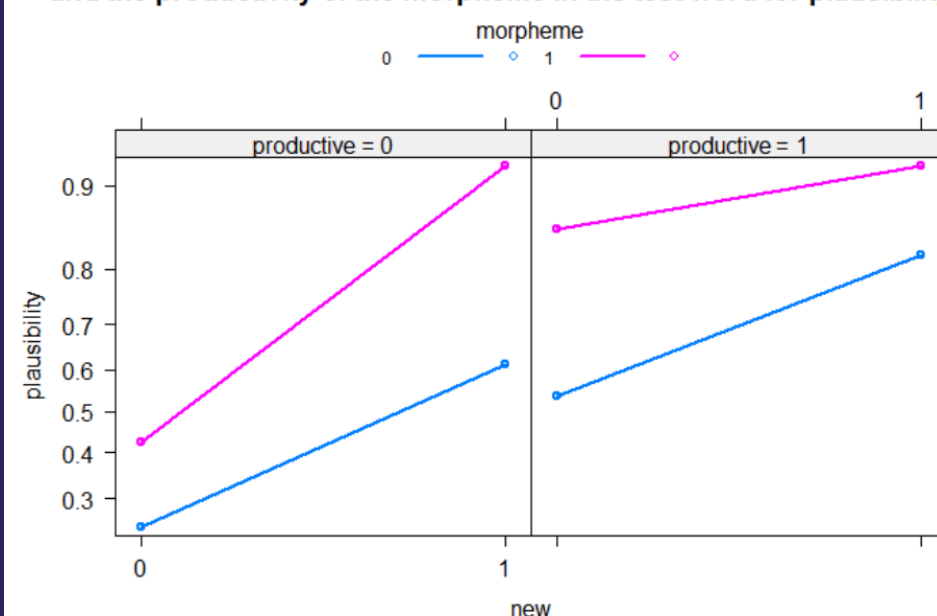
Discussion

- Adding contextual information = better performance;
–more productive suffixes = more frequent > more consistent form-meaning mappings [2], but bigger perplexity;
–less productive = more specific definitions;
- humans define new words differently;
- interaction between productivity and the word 'new': less strong for productive suffixes
–Why? productive suffixes defined in terms of hapax legomena; their roots = specialised meanings

- Note: limited sample of prompts, test items, human subjects, and languages.

violinous: Adjective, violin-like quality in sound or appearance, of music/of an object/of a composition.

2