gemini paper

index

the einleitung

inspired by the paper *Empirical evidence of Large Language Model's influence on human spoken communication Yakura et al. (2025)*, who indeed found (evidence) for GPT influenced human language after the introduction of chatGPT we tried to replicate the pipeline of building an AI vocabulary (gpt preferred lemmata) and compare frequencies of gpt-typical words across pre- and post chatGPT human language corpora. The first draft essai proves their hypothesis that LLM generated language manifests within human natural language.

preliminary

Our findings are still limited to a yet very small corpus of texts after the introduction of the google gemini chat agent to the public in 03/2024. In contrast to Yakura et al. (2025) and out of resources reasons we decided for gemini as basis for our AI generated vocabulary and for another text corpus (german bundestag plenary protocols, DIP (2026)) than youtube/podcast audio for the same reasons. That limits our post-AI corpus to a small timeframe between 03/2024 up to now. With expanding that corpus to a wider spectrum with including other sources we may harden our results.

hypothesis

following Yakura et al. (2025) we assumed that the consuming of LLM generated language influences the human production of language such that vocabulary typical for LLM output will be found with higher frequencies in human language corpora dating after chat agents introduction.

next methods

methods

snc

16062.1.2

data

our human langugae data consists of raw texts from german bundestag plenary protocols (DIP (2026)). the LLM corpus consists of model summaries of a first subset of these texts generated with the following prompt: Section .

corpus subsets

target	tokens
gemini	3957
human-pre	1514663
human-post	1458323

gemini prompt

```
[1] "System prompt: "
[2] "You are a member of german parliament. Prepare a summary of the text provided to presen
[3] "Text:"
```

computation

we first devised AI-typical lemmata in the model corpus which are distinctive for that corpus using a linear regression model (R, package lme4::glmer(): Bates et al. (2015)) that calculates a score for each lemma in the corpus, see Figure 1 and Figure 2.

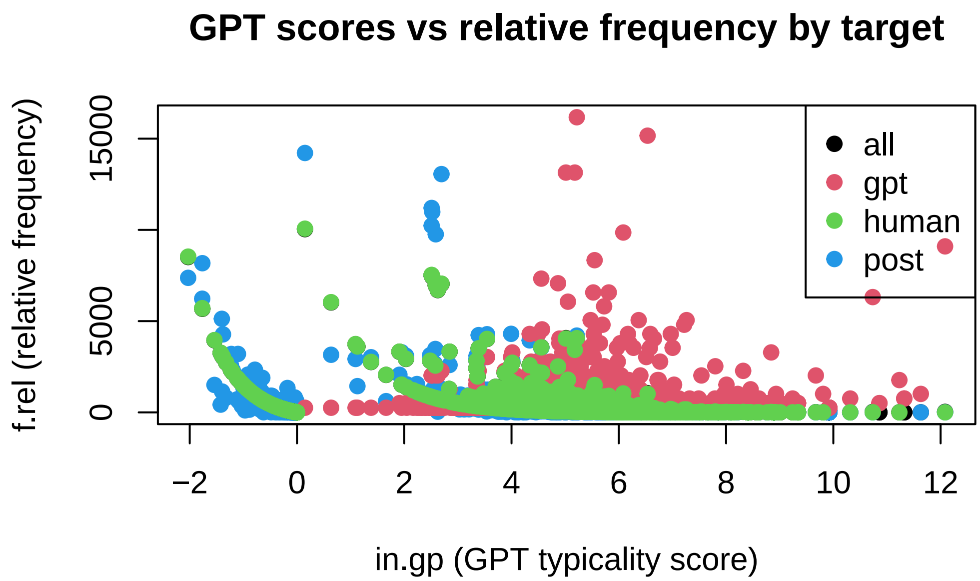

Figure 1: lemma gpt scores over targets

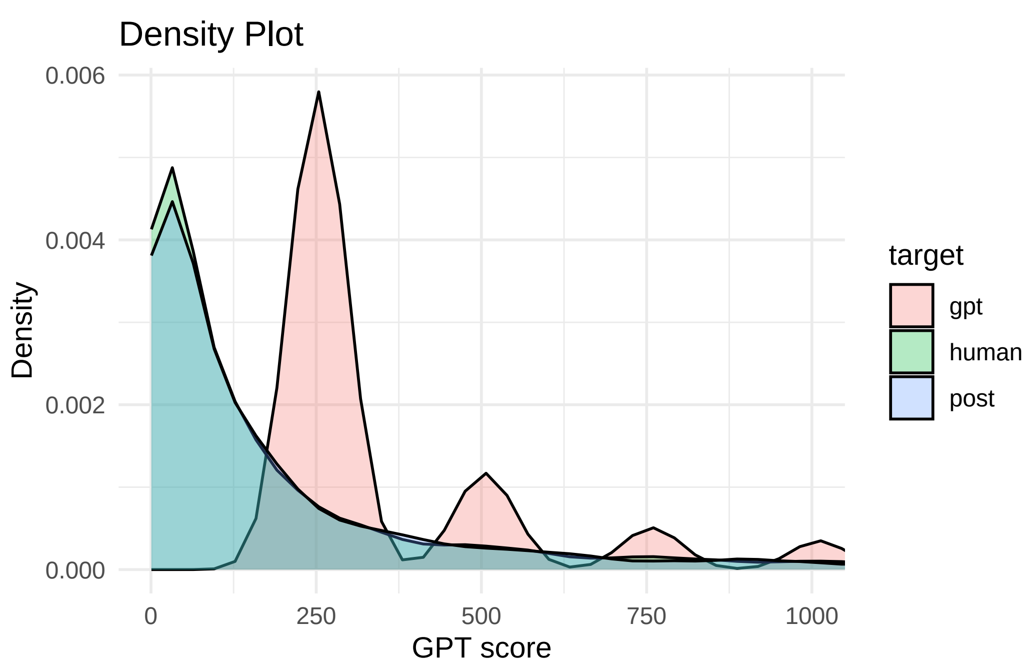

Figure 2: target gpt density

evaluation

basic descriptive

to first gather an insight, yet with simple descriptive stats comparing the raw frequencies of gpt-preferred lemmas in pre- and post-gemini onset we find that in the target corpus the occurences of these lemmas increase, only by small amount (see Table 2) and hard to visualise (see Figure 3). if these findings become relevant, we'll see in Section where we evaluate the frequencies with a linear regression model.

Table 2: GPT lemma frequencies (table) over target. (freq / Mtoken)

target	freq
human	0.4028
post	0.4287
DIFF:	0.0259

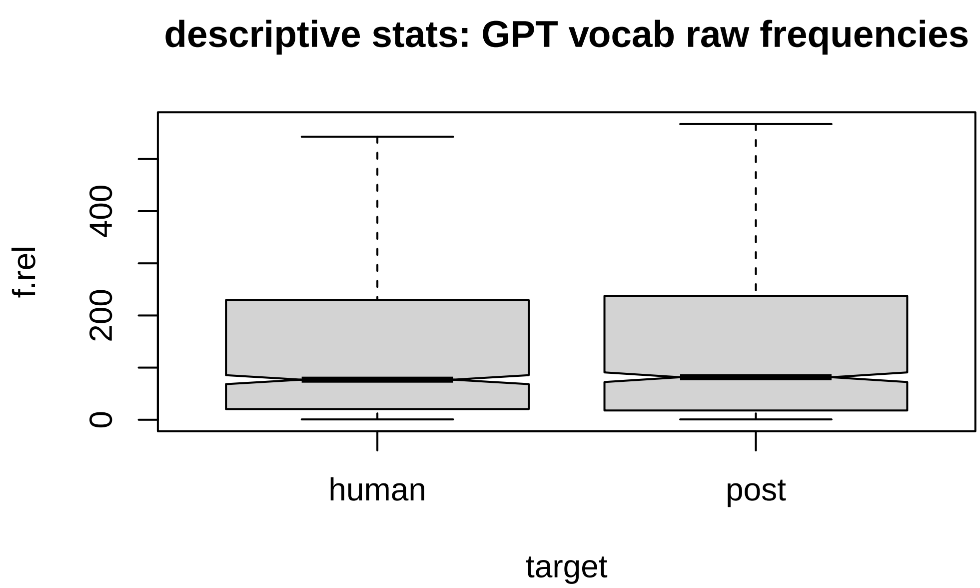

Figure 3: GPT lemma frequencies (boxplot) over target. (freq / Mtoken)

linear regression

to prove descriptive results, we compute the stability of the frequency variance for target and reference corpus with a linear regression model using R's lme4::lmer() function, cf. Bates et al. (2015). coefficents are printed below, where frequency are the relative lemma frequencies over corpus; target defines reference resp. target corpus[post-gpt] (human/post) and in.gpt as numerical variable representing the gpt-score of the corresponding lemma i.e. wether it scores high (positive values) or low (negative values) in terms of being preferredly used by the chat agent.

```
Loading required package: Matrix


Attaching package: 'lmerTest'


The following object is masked from 'package:lme4':

    lmer


The following object is masked from 'package:stats':

    step
```

basic (lm)

formula: `frequency.relative ~ target * in.gpt`

```
Call:
lm(formula = f.rel ~ target * in.gp, data = lmdf.c)

Residuals:
    Min      1Q  Median      3Q     Max 
 -409.3    -9.9    -9.8    -7.8 15521.0 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)       10.52309    0.58950  17.851  < 2e-16 ***
targetall          0.06232    0.83365   0.075   0.9404    
targetgpt        632.30872   14.50242  43.600  < 2e-16 ***
targetpost         0.95434    0.85908   1.111   0.2666    
```

```
in.gp             26.79998    0.75162  35.656  < 2e-16 ***
targetall:in.gp   -2.30237    1.03887  -2.216   0.0267 *
targetgpt:in.gp  -24.86712    2.60445  -9.548  < 2e-16 ***
targetpost:in.gp   4.24178    1.07704   3.938 8.21e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 162 on 223217 degrees of freedom
Multiple R-squared:  0.1106,    Adjusted R-squared:  0.1106
F-statistic:  3967 on 7 and 223217 DF,  p-value: < 2.2e-16
```

mixed effects model (lmer)

formula: `frequency.relative ~ target * in.gpt +(1|lemma)`

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: f.rel ~ target * in.gp + (1 | lemma)
   Data: lmdf.c

REML criterion at convergence: 2825867

Scaled residuals:
    Min      1Q  Median      3Q     Max
-60.813  -0.028  -0.026  -0.016 105.684

Random effects:
 Groups   Name        Variance Std.Dev.
 lemma    (Intercept) 10509    102.5
 Residual             11031    105.0
Number of obs: 223225, groups:  lemma, 110404

Fixed effects:
                   Estimate Std. Error         df t value Pr(>|t|)
(Intercept)       9.103e+00  5.182e-01  2.099e+05  17.568  < 2e-16 ***
targetall        -6.137e-02  5.403e-01  1.449e+05  -0.114   0.9096
targetgpt        -2.801e+02  1.063e+01  1.671e+05 -26.352  < 2e-16 ***
targetpost       -4.431e-01  6.366e-01  1.935e+05  -0.696   0.4864
in.gp             2.329e+01  6.705e-01  1.987e+05  34.738  < 2e-16 ***
targetall:in.gp   1.435e+00  6.783e-01  1.465e+05   2.116   0.0344 *
targetgpt:in.gp   1.268e+02  1.897e+00  1.666e+05  66.827  < 2e-16 ***
targetpost:in.gp  3.098e+00  7.051e-01  1.471e+05   4.393 1.12e-05 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) trgtll trgtgp trgtps in.gp  trgtl:. trgtg:.
targetall   -0.521
targetgpt   -0.029  0.027
targetpost  -0.583  0.424  0.027
in.gp       -0.117  0.062  0.034  0.069
trgtll:n.gp  0.064 -0.120 -0.033 -0.052 -0.537
trgtgpt:n.g  0.025 -0.023 -0.935 -0.024 -0.220  0.218
trgtpst:n.g  0.081 -0.059 -0.003 -0.140 -0.509  0.501   0.179
```

anova of mixed effects model

```
Analysis of Variance Table

Response: f.rel
                 Df     Sum Sq   Mean Sq  F value    Pr(>F)
target            3  622635042 207545014 7903.933 < 2.2e-16 ***
in.gp             1  102891443 102891443 3918.413 < 2.2e-16 ***
target:in.gp      3    3667275   1222425   46.554 < 2.2e-16 ***
Residuals    223217 5861332083     26258
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using Lme4." *Journal of Statistical Software* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01.

DIP. 2026. "DIP - Bundestagsprotokolle." Docs. *DIP - API*. Berlin. https://dip.bundestag.de/%C3%BCber-dip/hilfe/api#content.

Yakura, Hiromu, Ezequiel Lopez-Lopez, Levin Brinkmann, Ignacio Serna, Prateek Gupta, Ivan Soraperra, and Iyad Rahwan. 2025. "Empirical Evidence of Large Language Model's Influence on Human Spoken Communication." arXiv. https://doi.org/10.48550/arXiv.2409.01754.