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	3895
human-pre	1437497
human-post	1363467

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.

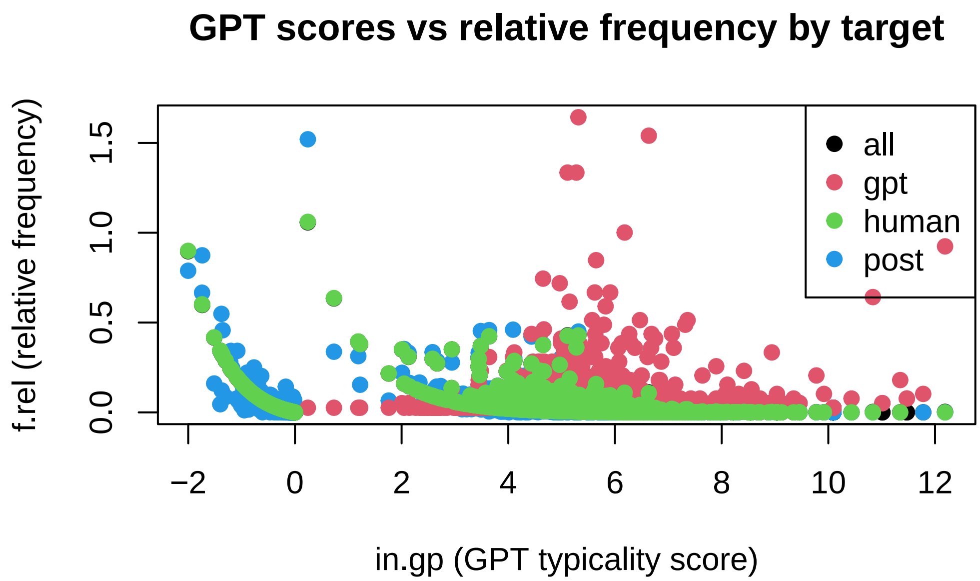

Figure 1: lemma gpt scores over targets

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 lemma 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	0.3708
post	0.3890
DIFF:	0.0182

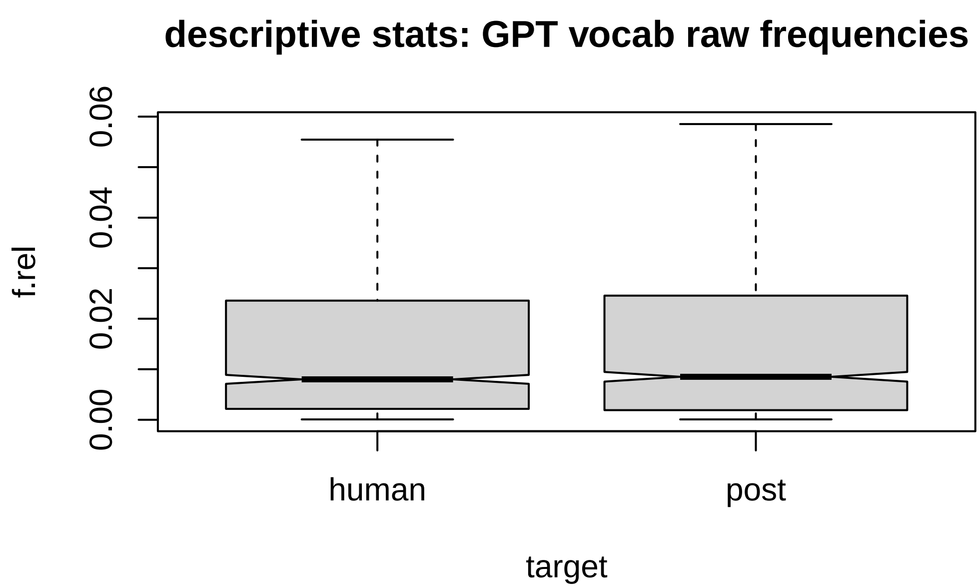

Figure 3: GPT lemma frequencies (boxplot) over target. (freq / Mtoken)

responsible lemmata

linear regression

to prove descriptive results, we compute the stability of the frequency increase for target- vs. 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 
-0.04285 -0.00099 -0.00099 -0.00077  1.57754 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)      1.061e-03  5.421e-05  19.570  < 2e-16 ***
targetall        5.864e-06  7.666e-05   0.077   0.9390    
```

```
targetgpt         6.128e-02  1.370e-03  44.733  < 2e-16 ***
targetpost        9.944e-05  7.899e-05   1.259   0.2081
in.gp             2.535e-03  6.791e-05  37.333  < 2e-16 ***
targetall:in.gp  -2.157e-04  9.387e-05  -2.298   0.0215 *
targetgpt:in.gp  -1.922e-03  2.404e-04  -7.996 1.30e-15 ***
targetpost:in.gp  3.859e-04  9.731e-05   3.965 7.33e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0149 on 223173 degrees of freedom
Multiple R-squared:  0.1278,    Adjusted R-squared:  0.1278
F-statistic:  4672 on 7 and 223173 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: -1319734

Scaled residuals:
    Min      1Q  Median      3Q     Max
-70.063  -0.031  -0.028  -0.017 115.436

Random effects:
 Groups   Name        Variance  Std.Dev.
 lemma    (Intercept) 8.870e-05 0.009418
 Residual             9.543e-05 0.009769
Number of obs: 223181, groups:  lemma, 110393

Fixed effects:
                   Estimate Std. Error        df t value Pr(>|t|)
(Intercept)       9.193e-04  4.795e-05  2.101e+05  19.171  < 2e-16 ***
targetall        -5.765e-06  5.026e-05  1.447e+05  -0.115   0.9087
targetgpt        -1.781e-02  1.015e-03  1.676e+05 -17.535  < 2e-16 ***
targetpost       -3.930e-05  5.911e-05  1.937e+05  -0.665   0.5061
in.gp             2.216e-03  6.094e-05  1.992e+05  36.359  < 2e-16 ***
targetall:in.gp   1.262e-04  6.200e-05  1.463e+05   2.035   0.0419 *
```

```
targetgpt:in.gp   1.096e-02  1.772e-04  1.672e+05  61.847  < 2e-16 ***
targetpost:in.gp  2.781e-04  6.444e-05  1.469e+05   4.315  1.6e-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   -0.524
targetgpt   -0.028  0.026
targetpost  -0.585  0.425  0.027
in.gp       -0.118  0.063  0.034  0.070
trgtll:n.gp  0.065 -0.121 -0.034 -0.053 -0.540
trgtgpt:n.g  0.025 -0.023 -0.938 -0.023 -0.218  0.214
trgtpst:n.g  0.082 -0.059 -0.003 -0.140 -0.512  0.501   0.175
```

anova of mixed effects model

```
Analysis of Variance Table

Response: f.rel
                 Df Sum Sq Mean Sq  F value    Pr(>F)
target            3  6.274 2.09120 9421.203 < 2.2e-16 ***
in.gp             1  0.961 0.96062 4327.761 < 2.2e-16 ***
target:in.gp      3  0.025 0.00822   37.038 < 2.2e-16 ***
Residuals    223173 49.537 0.00022
---
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.