Not that much power: Linguistic alignment is influenced more by low-level linguistic features rather than social power

Yang Xu and **Jeremy Cole** and **David Reitter**
College of Information Sciences and Technology
The Pennsylvania State University
innerfirexy@gmail.com, jrc436@psu.edu, reitter@psu.edu

PennState College of Information Sciences and Technology

Abstract

Does the social power status of people influence how others speak towards them? Previous studies said yes, but they overlooked the effect low-level linguistic features. We find that after controlling for low-level features (*utterance length*), the effect of power on alignment vanishes or is reversed.

Introduction

- The *social power* of interlocutors influences how they align (coordinate, accommodate) towards each other.
- Qualitative evidence: interviewers & interviewees (Willemyns et al., 1997), teachers & students (Jones et al., 1999) etc.
- Quantitative evidence: admins & non-admins in Wikipedia talkpages, judges & lawyers in supreme-court dialogues (Danescu-Niculescu-Mizil et al., 2012).
- **A widely accepted conclusion**: people align their language use more towards interlocutors of higher power (e.g., admins, judges) than those of lower power (e.g., non-admins, lawyers).
- **However**, previous studies overlooked the **low-level features** that could also affect alignment, e.g., lexical information density, syntactic surprisal, temporal clustering (Jaeger and Snider, 2008, 2013; Xu and Reitter, 2018; Mysln and Levy, 2016), which casts doubt on the conclusion.
- **Our work**: A two-step model analysis on how reliably social power affects alignment:
  - Step 1, a basic model to replicate Danescu-Niculescu-Mizil et al. (2012)'s findings.
  - Step 2, an extended model that includes *utterance length* on top of the basic model, aiming to examine whether the effect of social power still exists.

Experiment 1: Basic Model

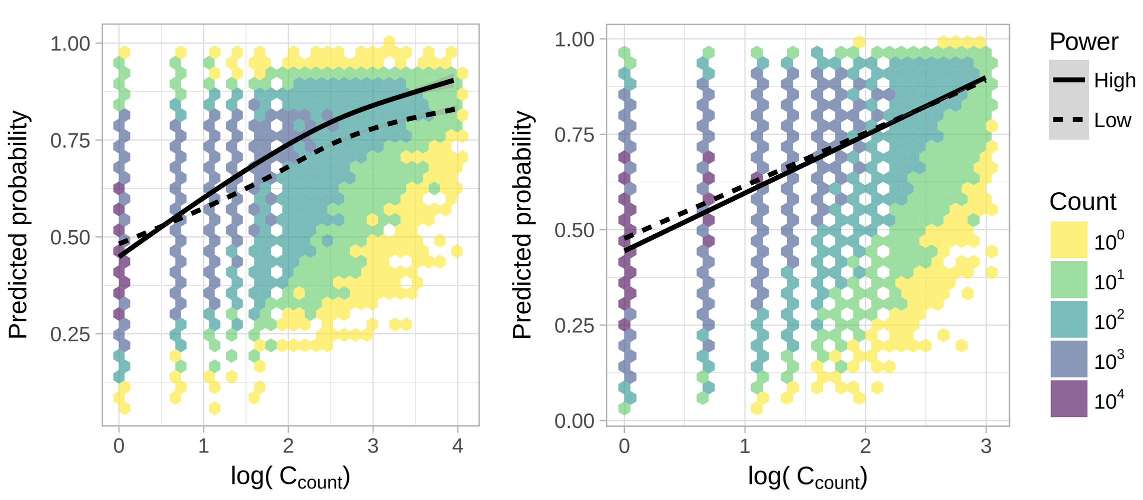
- Alignment is the impact of using certain linguistic elements in the preceding utterance (*prime*) on their chance to appear again in the following utterance (*target*).
- In the language of generalized linear models (GLM), we use the occurrence of linguistic markers in *target* as the response variable, and their occurrence in *prime* as the predictor.

$$\begin{aligned}\mathrm{logit}(m) &= \ln\frac{p(m \text{ in } target)}{p(m \text{ not in } target)}\\ &= \beta_0 + \beta_1 C_{\text{count}} + \beta_2 C_{\text{power}}\\ &+ \beta_3 \mathbf{C_{count} * C_{power}}\end{aligned} \tag{1}$$

- Here, C_{count} is the number of marker m in *prime*. C_{power} is a binary predictor indicating the power status of *prime* speaker (*high* vs. *low*).
- A linguistic marker m is one of the 14 LIWC[1] categories: adverbs, articles, auxiliary verbs, certainty, conjunctions, discrepancy, exclusive, inclusive, impersonal pronouns, negations, personal pronouns, prepositions, quantifiers, and tentativeness.
- Datasets: Wikipedia talk-page corpus (*Wiki*) and a corpus of United States supreme court conversations (*SC*) (compiled by Danescu-Niculescu-Mizil et al. (2012)).
- **Results**: z scores of β_3 are shown in Table 1. Slopes of C_{count} are visualized in Figure 1.

Marker	$C_{\text{count}} * C_{\text{power}}$ SC	Wiki
adv	**6.16*****	-0.40
art	**4.60*****	1.27
auxv	**5.81*****	-0.83
certain	**1.94**†	**2.84****
conj	**6.79*****	0.39
discrep	**8.03*****	0.25
excl	**2.94****	**2.16***
incl	**5.24*****	**2.15***
ipron	**10.22*****	**1.90**†
negate	**5.49*****	**3.11****
ppron	1.29	-1.13
prep	**6.87*****	-0.19
quant	**4.14*****	-0.04
tentat	**4.52*****	-0.78

Table 1: Wald's z-score and significance level (*** for $p < 0.001$, ** for $p < 0.01$, * for $p < 0.05$, and † for $0.05 < p < 0.1$) of the interaction term, $C_{\text{count}} * C_{\text{power}}$.



a: Supreme Court b: Wikipedia

Figure 1: The predicted probability against C_{count}, grouped by C_{power}. Divergent slopes indicate significant interaction: high power speakers have larger slopes than low power ones.

- Therefore, we have replicated the previous finding from Danescu-Niculescu-Mizil et al. (2012): significant β_3 of $C_{\text{count}} * C_{\text{power}}$ indicates that the β of C_{count} depends on C_{power}, i.e., the strength of alignment varies with the power levels of speakers (*high* vs. *low*).
- **However**, this affirmative finding is **not** safe, because only one predictor, C_{power}, is included in the model, which we will show in Experiment 2.

Experiment 2: Extended Model

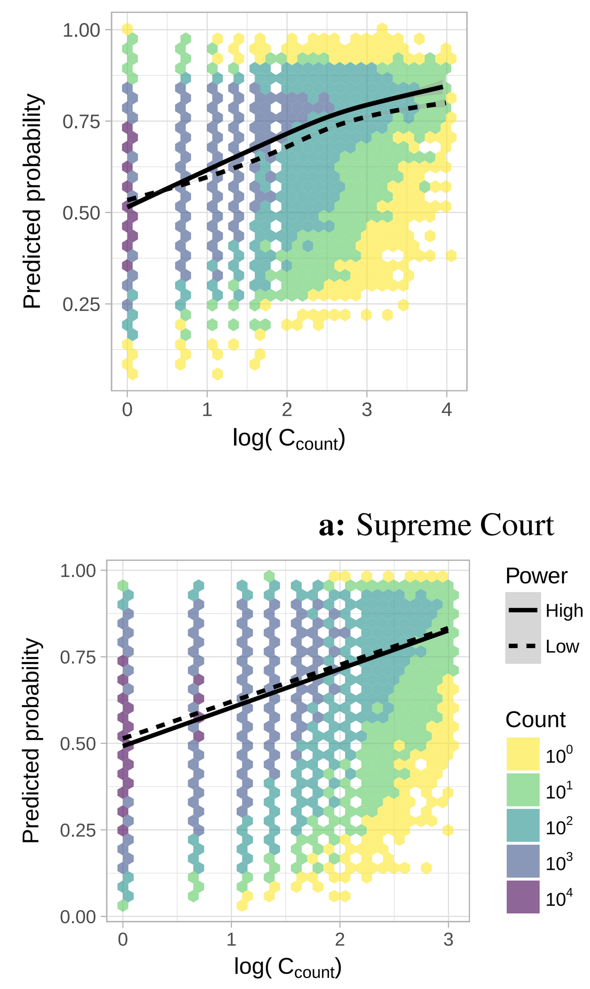
- Does the interaction term $C_{\text{count}} * C_{\text{power}}$ remain significant *after* including other predictors that represent low-level linguistic features?
- We add **utterance length** (number of words) as an additional predictor to the model, C_{pLen}.

$$\begin{aligned}\mathrm{logit}(m) &= \ln\frac{p(m \text{ in } target)}{p(m \text{ not in } target)}\\ &= \beta_0 + \beta_1 C_{\text{count}} + \beta_2 C_{\text{power}} + \beta_3 C_{\text{pLen}}\\ &+ \boldsymbol{\beta_4 C_{count} * C_{power}}\\ &+ \beta_5 C_{\text{count}} * C_{\text{pLen}}\\ &+ \beta_6 C_{\text{power}} * C_{\text{pLen}}\\ &+ \beta_7 C_{\text{count}} * C_{\text{power}} * C_{\text{pLen}}\end{aligned} \tag{2}$$

- Our goal: to examine whether β_4 remains significant and in same direction as β_3 in Experiment 1.
- **Results**: Full model coefficients are shown in Table 2.

Table 2: Model coefficients of all terms in Equation 2. Notice that the β_4 of $C_{\text{count}} * C_{\text{power}}$ is **negative** in SC and **non-significant** in Wiki

Corpus	Predictor	β	z
SC	Intercept	0.360	2.40*
	C_{count}	0.213	26.92***
	C_{power}	-0.060	-3.39***
	C_{pLen}	0.080	13.03***
	$\mathbf{C_{count} * C_{power}}$	**-0.103**	**-9.95*****
	$C_{\text{count}} * C_{\text{pLen}}$	-0.066	-15.35***
	$C_{\text{power}} * C_{\text{pLen}}$	0.231	25.25***
	$C_{\text{count}} * C_{\text{power}} * C_{\text{pLen}}$	0.036	4.79***
Wiki	Intercept	0.330	1.40
	C_{count}	0.149	31.11***
	C_{power}	-0.074	-10.52***
	C_{pLen}	0.179	40.80***
	$\mathbf{C_{count} * C_{power}}$	**0.001**	**0.14**
	$C_{\text{count}} * C_{\text{pLen}}$	0.022	6.13***
	$C_{\text{power}} * C_{\text{pLen}}$	0.042	5.52***
	$C_{\text{count}} * C_{\text{power}} * C_{\text{pLen}}$	-0.010	-1.61

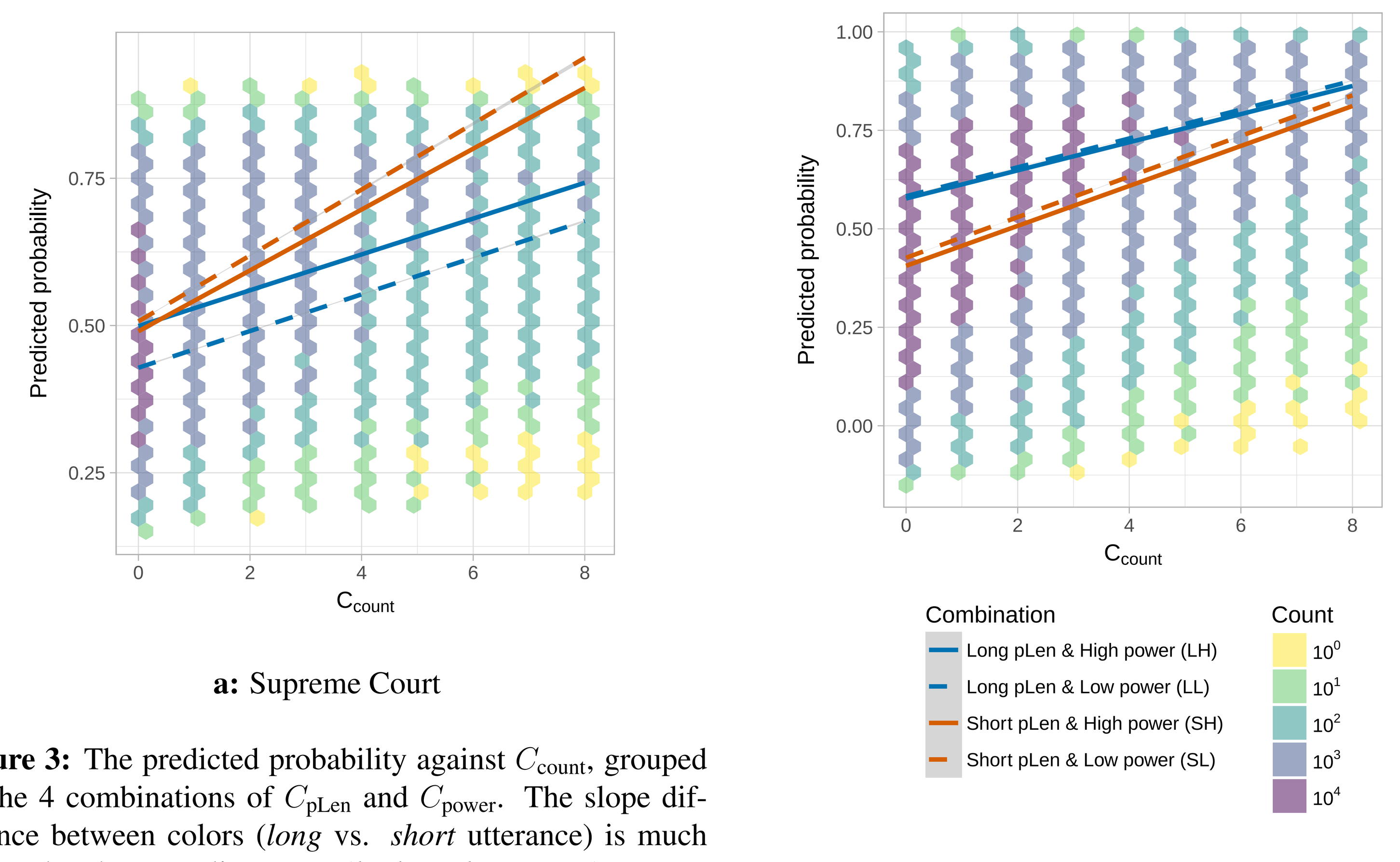


a: Supreme Court

b: Wikipedia

Figure 2: The predicted probability against C_{count}, grouped by C_{power}. Notice that the differences in slopes (between *high* power and *low* power) are much smaller than Figure 1.

- **Surprisingly**, the coefficient of $C_{\text{count}} * C_{\text{power}}$ is negative in SC, and non-significant in Wiki, which is inconsistent with the positive coefficients in Experiment 1.
- Further illustration of how the interaction $C_{\text{count}} * C_{\text{power}}$ diminishes after including C_{pLen} to the model: cluster C_{pLen} to two discrete values, *long* and *short*, and then plot the regression lines grouped by the combination of C_{pLen} and C_{power} (shown in Figure 3).



a: Supreme Court

b: Wikipedia

Figure 3: The predicted probability against C_{count}, grouped by the 4 combinations of C_{pLen} and C_{power}. The slope difference between colors (*long* vs. *short* utterance) is much bigger than between line-types (*high* vs. *low* power).

- Thus, C_{pLen} is a more determinant factor of alignment than C_{power}.

Discussion and Conclusions

- Our findings suggest that the previously reported effect of power on linguistic alignment is not reliable. Instead, alignment is more sensitive to certain low-level features (e.g., utterance length).
- We do not deny the existence of accommodation caused by social perception, but we want to emphasize the difference between the priming-induced alignment and the intentional accommodation.
- The dynamics of LIWC categories in probability space is more likely to be a case of automatic alignment, rather than accommodation. Therefore, we suggest future work on social aspects of language use should consider higher levels of linguistic elements.

Acknowledgements

This work has been funded by the National Science Foundation (IIS-1459300 and BCS- 1457992).

[1] Linguistic Inquiry and Word Count, https://liwc.wpengine.com/