Linguistic alignment is affected more by lexical surprisal rather than social power

Yang Xu[1], Jeremy Cole[2,3], and David Reitter[2,4]

[1]Department of Computer Science, San Diego State University
[2]College of Information Sciences and Technology, The Pennsylvania State University
[3]Google Inc., Mountain View, CA
[4]Google Inc., New York, NY
yxu4@sdsu.edu, jrcole@psu.edu, and reitter@psu.edu

1 Introduction

Social power has been found to affect the linguistic alignment between dialogue interlocutors (Giles, 2008; Willemyns et al., 1997; Jones et al., 1999). A common finding from some recent data-driven studies is that interlocutors of lower power positions tend to receive more alignment than those of higher power (Danescu-Niculescu-Mizil et al., 2012).

However, this finding is surprising from a psycholinguistic perspective, because the mutual alignment between interlocutors of a natural dialog is considered to be due to the autonomous and low level priming process (Pickering and Garrod, 2004). A wealth of studies on alignment at syntactic levels, i.e., *structural priming*, have shown that alignment is sensitive to various features of the linguistic properties of the utterance, such as syntactic and lexical surprisal (Jaeger and Snider, 2008, 2013; Xu and Reitter, 2018), temporal clustering (Myslín and Levy, 2016) etc.

Then why, or under what mechanisms, would alignment be affected by the relatively high-level social perceptions of power? Could it be the case that the observed effect of power on alignment is actually due to some other factors in language per se, such as the temporal clustering or the surprisal of linguistic elements as mentioned above? If after ruling out all the other factors, the effect of power still exists, how large is the effect? Is it significant enough to cause the difference that can be captured by computational measures of alignment? Answering these questions is important for us to resolve the confusions among different research approaches of linguistic alignment, and to obtain a better understanding of language production in general.

In order to answer the above questions, we setup a series of model analysis. First, we use a baseline model to replicate the previous finding. Then we include additional predictors to examine whether the effect of power on alignment is reliable. To clarify, our goal is to explore alternative explanations before quickly reaching conclusions about the effect of social power on alignment.

2 Method and Results

2.1 Data and Processing

Two datasets are used: Wikipedia talk-page corpus (*Wiki*) and a corpus of United States supreme court conversations (*SC*), both of which are used in Danescu-Niculescu-Mizil et al. (2012). Wiki is a collection of conversations between Wikipedia editors [1]. SC is a collection of conversations from the U.S. Supreme Court Oral Arguments[2], from 204 cases involving 11 Justices and 311 other participants (lawyers or amici curiae).

Two distinct social power statues exist in both corpora. In Wiki, admins are of higher power than editors, while in SC, judges have higher power than lawyers.

A conversation is a sequence of utterances, $\{u_i\}(i = 1, 2, \ldots, N)$. u_i and u_{i+1} are from different speakers because of turn taking. We use a window of 2 to scan through the conversation, generating a sequence of adjacent utterance pairs, $\{\langle prime_i, target_i\rangle\}(i = 1, 2 \ldots N - 1)$. we are interested in the alignment between those pairs.

2.2 Characterize alignment with GLM

We formulate alignment as the impact of using certain linguistic markers in the preceding utterance (prime) on their chance to appear again in the following utterance (target). The linguistic mark-

[1] http://en.wikipedia.org/wiki/Wikipedia:Talk_page_guidelines
[2] http://www.supremecourt.gov/oral_arguments/

Proceedings of the Society for Computation in Linguistics (SCiL) 2019, pages 349-352.
New York City, New York, January 3-6, 2019

Table 1: Examples of the 14 linguistic markers used in this study.

LIWC category	Examples
Adverbs (*adv*)	*actually*, *totally*
Articles (*art*)	*a*, *the*
Auxiliary verbs (*auxv*)	*can*, *could*
Certainty (*certain*)	*always*, *never*
Conjunctions (*conj*)	*but*, *and*
Discrepancy (*discrep*)	*should*, *would*
Exclusive (*excl*)	*without*, *except*
Inclusive (*incl*)	*with*, *along*
Impersonal pronouns (*ipron*)	*it*, *itself*
Negation (*negate*)	*not*, *never*
Personal pronouns (*ppron*)	*I*, *you*, *we*
Prepositions (*prep*)	*to*, *in*, *by*
Quantifiers (*quant*)	*few*, *many*
Tentativeness (*tentat*)	*perhaps*, *maybe*

ers examined here are 14 LIWC categories (Pennebaker et al., 2001) as shown in Table 1.

We use generalized linear models (GLMs), in which the occurrence of linguistic markers (binary) in *target* as the response variable; the predictor is their occurrence (frequency count) in *prime*. Thus alignment is characterized by the β coefficient of the predictor. Larger (positive) β indicates stronger influence from prime on target, i.e., stronger alignment.

2.3 Baseline model

To replicate Danescu-Niculescu-Mizil et al. (2012)'s results, we set a baseline model with two predictors, marker counts in prime, C_{count} (numeric), and the power status of prime speaker, C_{power} (binary, high vs. low). The formula is:

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

Indeed our results from the baseline model are consistent with previous findings: β_3 is significant and *positive* in 13 out of 14 markers ($p < .001$). Because $\beta_3 > 0$ holds, for high power speaker ($C_{power} = 1$), the coefficient of C_{count}, $\beta_1 + \beta_3$ is larger than β_1, which is the coefficient of C_{count} for low power speakers.

No collinearity is found between C_{count} and C_{power} (Pearson correlation $r < 0.2$).

2.4 Baseline + Surprisal

We want to examine whether the interaction term $C_{\text{count}} * C_{power}$ still remains significant and positive after the surprisal of utterances is taken into account. We include a third predictor C_{pSurp} into the model, which is the total amount of lexical surprisal of the utterance. The formula now becomes:

$$\begin{aligned}\text{logit}(\textit{marker}) = \beta_0 &+ \beta_1 C_{\text{count}} + \beta_2 C_{\text{power}} + \beta_3 \boldsymbol{C}_{\mathbf{pSurp}} \\ &+ \beta_4 C_{\text{count}} * C_{\text{power}} \\ &+ \beta_5 C_{\text{count}} * \boldsymbol{C}_{\mathbf{pSurp}} \\ &+ \beta_6 C_{\text{power}} * \boldsymbol{C}_{\mathbf{pSurp}} \\ &+ \beta_7 C_{\text{count}} * C_{\text{power}} * \boldsymbol{C}_{\mathbf{pSurp}}\end{aligned} \tag{2}$$

Here surprisal is defined as the total negative log probability of all words within the utterance: $\log \prod_{w_k \in u} p(w_k | w_1 w_2 \dots w_{k-1})$, where the conditional probabilities are estimated by a trigram language model.

Motivation

Surprisal has been found to be an important factor that affects a borad range of aspects in language comprehension and production. Jaeger and Snider (2008) finds that structural priming in spontaneous dialogue is sensitive to the surprisal of structures: less common structures have stronger priming effect. It also has been found that variation of sentence complexity (Xu and Reitter, 2016) and information density (Xu and Reitter, 2018) relate closely to lexical alignment. Therefore, it makes sense to expect that the alignment of LIWC markers is also affected by surprisal.

Results

Having C_{pSurp} in the model, our results show that β_4 becomes non-significant in Wiki ($\beta_4 = 0.04$, $p > .05$), and even negative in SC ($\beta_4 = -0.098$, $p < .001$). This is in contrast to the positive β_3 of the baseline model. It indicates that the previously observed effect of power on alignment is actually dependent on the presence of predictor C_{pSurp}. Also, the coefficients of the three-way interaction term $C_{\text{count}} * C_{\text{power}} * C_{\text{pSurp}}$ are significant, which indicates that the interaction between C_{power} and C_{count} will be influenced by the presence of C_{pSurp} in the model.

No collinearity is found between C_{power} and other predictors: Pearson correlation $r < 0.2$; Variance inflation factor (VIF) is low (< 2.0).

2.5 Baseline + Clustering

Here we examine whether another low level linguistic feature, the clustering of markers, can draw similar results as surprisal.

Motivation

How some certain language structures cluster in

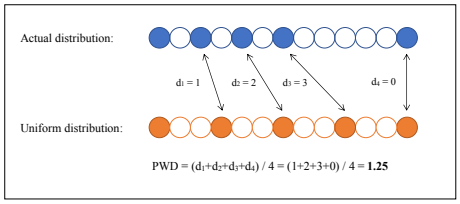


Figure 1: Operational definition of clustering.

time affects the comprehension of them. Myslín and Levy (2016) showed that sentence comprehension is faster when the same syntactic structure clusters in time in prior experience than when it is evenly spaced in time. The the alignment studied here is due to comprehension-to-production priming, it is reasonable to anticipate that the alignment of LIWC markers may also be influenced by their clustering patterns.

Operational definition

To measure the degree to which linguistic markers cluster, we use the point-wise distance (PWD) between two sequences: the first one represents the actual distribution of a marker within the utterance; the second one is a uniform distributed sequence representing the ideal non-clustering case. A demonstration of computation is shown in Section 2.5. Larger PWD values indicate stronger clustering property.

Having computed PWD for each utterance, we use it as a new predictor C_{pCls} ("Cls" stands for clustering) to replace C_{pSurp} in Equation (2), resulting in a new model:

$$\begin{aligned}\text{logit}(\textit{marker}) = \beta_0 &+ \beta_1 C_{\text{count}} + \beta_2 C_{\text{power}} + \beta_3 \boldsymbol{C_{\text{pCls}}} \\ &+ \beta_4 C_{\text{count}} * C_{\text{power}} \\ &+ \beta_5 C_{\text{count}} * \boldsymbol{C_{\text{pCls}}} \\ &+ \beta_6 C_{\text{power}} * \boldsymbol{C_{\text{pCls}}} \\ &+ \beta_7 C_{\text{count}} * C_{\text{power}} * \boldsymbol{C_{\text{pCls}}}\end{aligned} \tag{3}$$

Similarly, we want to examine whether β_4 the coefficient of $C_{\text{count}} * C_{\text{power}}$ in Equation (3) is significant and positive.

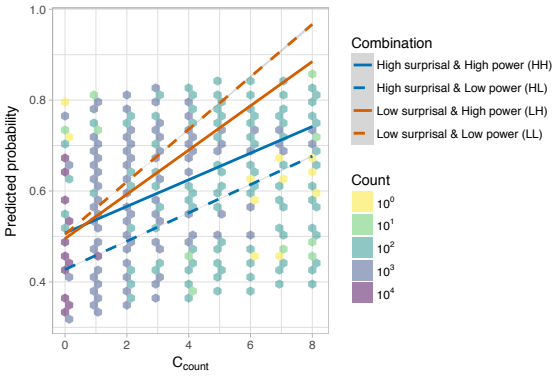
Results

Again, β_4 is negative in SC ($\beta_4 = 0.061$, $p < .01$), but it remains positive in Wiki ($\beta = 0.059$, $p < .01$). It indicates that the influence of clustering is weaker than surprisal.

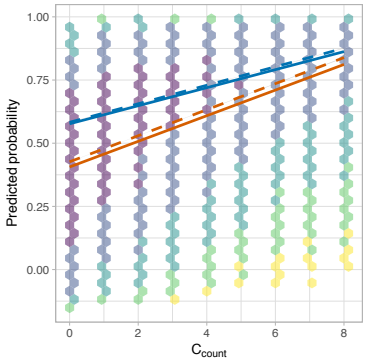
2.6 Model comparison

We compare the quality of the three models: Baseline, Baseline+Surprisal, and Baseline+Clustering, and find that Baseline+Surprisal has the lowest AIC score, which means that it is the best fit. Also, Baseline has the highest AIC score, meaning that the inclusion of surprisal and clustering is non-trivial.

Besides, we examine the model of using utterance length instead of surprisal as the predictor. It turns out that this model also diminishes or reverses the interaction effect of power, but it explains less variance in data than surprisal. Considering that length is a coarser estimation of information than surprisal, we do not report that part of results here.



(a) Supremecourt



(b) Wikipedia

Figure 2: Predicted probability of words appearing in following utterance against the number of same words in preceding utterance, grouped by four combinations: high and low surprisal (color), high and low power (line type). Hexagon indicates the number of data points within that region.

2.7 Interaction analysis

To better illustrate how the interaction $C_{\text{power}} * C_{\text{count}}$ diminishes after introducing C_{pSurp} into the model, we visualize the interaction effects in Figure 2. In details, we cluster C_{pSurp} into two groups, low and high, and then observe how the amount of

priming changes with C_{count} , with respect to different combinations of C_{power} and C_{pSurp}. This is a common practice to interpret linear models that consist of three-way interactions (Houslay, 2014).

Figure 2 intuitively shows that C_{pSurp} is a more determinant predictor than C_{power}, because the differences in slopes are larger between colors (high vs. low surprisal) than between power (high vs. low power).

3 Discussion and Conclusions

We have presented three experiments to replicate the previous finding on social power, but then to show that the effect of power is unreliable when also considering linguistic features. Instead, we consistently align towards the language that share certain low-level features, especially those of higher suprisal. We call for the consideration of taking into account wider range of factors for future studies on social factors of language use, especially those low level yet straight forward cognitive factors that has strong predictability of human language.

We are not denying the existence of accommodation caused by the social distance between interlocutors. However, we want to stress the difference between the priming-induced alignment at lower linguistic levels and the intentional accommodation that is caused by higher-level perception of social power. The latter should be a relatively stable effect that is independent of the low-level linguistic features. In particular, our findings suggest that the probability change of LIWC categories is more likely to be a case of automatic alignment, rather than an intentional accommodation, because it is better explained by lower-level linguistic features like surprisal.

Acknowledgement

This work was funded by the National Science Foundation (IIS-1459300 and BCS-1457992). It was presented as a poster in the 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018, at Melbourne, Australia) and published in the proceedings.

References

Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang, and Jon Kleinberg. 2012. Echoes of power: Language effects and power differences in social interaction. In *Proceedings of the 21st International Conference on World Wide Web*, pages 699–708, Lyon, France. ACM.

Howard Giles. 2008. *Communication accommodation theory*. Sage.

Thomas M. Houslay. 2014. Understanding 3-way interactions between continuous variables. https://tomhouslay.com/2014/03/21/understanding-3-way-interactions-\between-continuous-variables/.

T Florian Jaeger and Neal Snider. 2008. Implicit learning and syntactic persistence: Surprisal and cumulativity. In *Proceedings of the cognitive science society conference*, pages 1061–1066.

T Florian Jaeger and Neal E Snider. 2013. Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, 127(1):57–83.

Elizabeth Jones, Cynthia Gallois, Victor Callan, and Michelle Barker. 1999. Strategies of accommodation: Development of a coding system for conversational interaction. *Journal of Language and Social Psychology*, 18(2):123–151.

Mark Myslín and Roger Levy. 2016. Comprehension priming as rational expectation for repetition: Evidence from syntactic processing. *Cognition*, 147:29–56.

James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, 71:2001.

Martin J Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(02):169–190.

Michael Willemyns, Cynthia Gallois, Victor Callan, and J Pittam. 1997. Accent accommodation in the employment interview. *Journal of Language and Social Psychology*, 15(1):3–22.

Yang Xu and David Reitter. 2016. Convergence of syntactic complexity in conversation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 443–448, Berlin, Germany. Association for Computational Linguistics.

Yang Xu and David Reitter. 2018. Information density converges in dialogue: Towards an information-theoretic model. *Cognition*, 170:147–163.