



IRLE WORKING PAPER  
#102-17  
April 2017

## Alignment at Work: Using Language to Distinguish the Internalization and Self-Regulation Components of Cultural Fit in Organizations

Gabriel Doyle, Sameer B. Srivastava, Amir Goldberg, and Michael C. Frank

Cite as: Gabriel Doyle, Sameer B. Srivastava, Amir Goldberg, and Michael C. Frank. (2017). "Alignment at Work: Using Language to Distinguish the Internalization and Self-Regulation Components of Cultural Fit in Organizations". IRLE Working Paper No. 102-17.  
<http://irle.berkeley.edu/files/2017/Alignment-At-Work.pdf>

# Alignment at Work: Using Language to Distinguish the Internalization and Self-Regulation Components of Cultural Fit in Organizations

**Gabriel Doyle**

Department of Psychology  
Stanford University  
gdoyle@stanford.edu

**Amir Goldberg**

Graduate School of Business  
Stanford University  
amirgo@stanford.edu

**Sameer B. Srivastava**

Haas School of Business  
UC Berkeley  
srivastava@haas.berkeley.edu

**Michael C. Frank**

Department of Psychology  
Stanford University  
mcfrank@stanford.edu

## Abstract

Cultural fit is widely believed to affect the success of individuals and the groups to which they belong. Yet it remains an elusive, poorly measured construct. Recent research draws on computational linguistics to measure cultural fit but overlooks asymmetries in cultural adaptation. By contrast, we develop a directed, dynamic measure of cultural fit based on linguistic alignment, which estimates the influence of one person’s word use on another’s and distinguishes between two enculturation mechanisms: internalization and self-regulation. We use this measure to trace employees’ enculturation trajectories over a large, multi-year corpus of corporate emails and find that patterns of alignment in the first six months of employment are predictive of individuals downstream outcomes, especially involuntary exit. Further predictive analyses suggest referential alignment plays an overlooked role in linguistic alignment.

## 1 Introduction

Entering a new group is rarely easy. Adjusting to unfamiliar behavioral norms and donning a new identity can be cognitively and emotionally taxing, and failure to do so can lead to exclusion. But successful enculturation to the group often yields significant rewards, especially in organizational contexts. Fitting in has been tied to positive career outcomes such as faster time-to-promotion, higher performance ratings, and reduced risk of being fired (O’Reilly et al., 1991; Goldberg et al., 2016).

A major challenge for enculturation research is distinguishing between *internalization* and *self-*

*regulation*. Inward-looking internalization involves identifying as a group member and accepting group norms, while outward-looking self-regulation entails deciphering the group’s normative code and adjusting one’s behavior to comply. Existing approaches, which generally rely on self-reports, are subject to self-report bias and typically yield only static snapshots of this process. Recent computational approaches that use language as a behavioral signature of group integration uncover dynamic traces of enculturation but cannot distinguish internalization and self-regulation.

To overcome these limitations, we introduce a dynamic measure of *directed* linguistic accommodation between a newcomer and existing group members. Our approach differentiates between an individual’s (1) *base rate* of word use and (2) *linguistic alignment* to interlocutors. The former corresponds to internalization of the group’s linguistic norms, whereas the latter reflects the capacity to regulate one’s language in response to peers’ language use. We apply this language model to a corpus of internal email communications and personnel records, spanning a seven-year period, from a mid-sized technology firm. We show that changes in base rates and alignment, especially with respect to pronoun use, are consistent with successful assimilation into a group and can predict eventual employment outcomes—continued employment, involuntary exit, or voluntary exit—at levels above chance. We use this predictive problem to investigate the nature of linguistic alignment. Our results suggest that the common formulation of alignment as a lexical-level phenomenon is incomplete.

## 2 Linguistic Alignment and Group Fit

**Linguistic alignment** Linguistic alignment is the tendency to use the same or similar words

as one's conversational partner. Alignment is an instance of a widespread and socially important human behavior: *communication accommodation*, the tendency of two interacting people to nonconsciously adopt similar behaviors. Evidence of accommodation appears in many behavioral dimensions, including gestures, postures, speech rate, self-disclosure, and language or dialect choice (see Giles et al. (1991) for a review). More accommodating people are rated by their interlocutors as more intelligible, attractive, and cooperative (Feldman, 1968; Ireland et al., 2011; Triandis, 1960). These perceptions have material consequences—for example, high accommodation requests are more likely to be fulfilled, and pairs who accommodate more in how they express uncertainty perform better in lab-based tasks (Buller and Aune, 1988; Fusaroli et al., 2012).

Although accommodation is ubiquitous, individuals vary in their levels of accommodation in ways that are socially informative. Notably, more powerful people are accommodated more strongly in many settings, including trials (Gnisci, 2005), online forums (Danescu-Niculescu-Mizil et al., 2012), and Twitter (Doyle et al., 2016). Most relevant for this work, speakers may increase their accommodation to signal camaraderie or decrease it to differentiate from the group. For example, Bourhis and Giles (1977) found that Welsh English speakers increased their use of the Welsh accent and language in response to an English speaker who dismissed it.

### Person-group fit and linguistic alignment

These findings suggest that linguistic alignment is a useful avenue for studying how people assimilate into a group. Whereas traditional approaches to studying person-group fit rely on self-reports that are subject to self-report bias and cannot feasibly be collected with high granularity across many points in time, recent studies have proposed language-based measures as a means to tracing the dynamics of person-group fit without having to rely on self-reports. Building on Danescu-Niculescu-Mizil et al. (2013)'s research into language use similarities as a proxy for social distance between individuals, Srivastava et al. (forthcoming) and Goldberg et al. (2016) developed a measure of cultural fit based on the similarity in linguistic style between individuals and their colleagues in an organization. Their time-varying measure highlights linguistic compatibility as an

important facet of cultural fit and reveals distinct trajectories of enculturation for employees with different career outcomes.

While this approach can help uncover the dynamics and consequences of an individual's fit with her colleagues in an organization, it cannot disentangle the underlying reasons for this alignment. For two primary reasons, it cannot distinguish between fit that arises from internalization and fit produced by self-regulation. First, Goldberg et al. (2016) and Srivastava et al. (forthcoming) define fit using a symmetric measure, the Jensen-Shannon divergence, which does not take into account the direction of alignment. Yet the distinction between an individual adapting to peers versus peers adapting to the individual would appear to be consequential. Second, this prior work considers fit across a wide range of linguistic categories but does not interrogate the role of particular categories, such as pronouns, that can be especially informative about enculturation. For example, a person's base rate use of the first-person singular (*I*) or plural (*we*) might indicate the degree of group identity internalization, whereas adjustment to *we* usage in response to others' use of the pronoun might reveal the degree of self-regulation to the group's normative expectations.

**Modeling fit with WHAM** To address these limitations, we build upon and extend the WHAM alignment framework (Doyle and Frank, 2016) to analyze the dynamics of internalization and self-regulation using the complete corpus of email communications and personnel records from a mid-sized technology company over a seven-year period. WHAM uses a conditional measure of alignment, separating overall homophily (unconditional similarity in people's language use, driven by *internalized* similarity) from in-the-moment adaptation (adjusting to another's usage, corresponding to *self-regulation*). WHAM also provides a directed measure of alignment, in that it estimates a *replier's* adaptation to the other conversational participant separately from the participant's adaptation to the replier.

**Level(s) of alignment** The convention within linguistic alignment research, dating back to early work on Linguistic Style Matching (Niederhoffer and Pennebaker, 2002), is to look at lexical alignment: the repetition of the same or similar words across conversation participants. From a commu-

nication accommodation standpoint, this is justified by assuming that one’s choice of words represents a stylistic signal that is partially independent of the meaning one intends to express—similar to the accommodation on paralinguistic signals discussed above. The success of previous linguistic alignment research shows that this is valid.

However, words are difficult to divorce from their meanings, and sometimes repeating a word conflicts with repeating its referent. In particular, pronouns often refer to different people depending on who uses the pronoun. While there is evidence that one person using a first-person singular pronoun increases the likelihood that her conversation partner will as well (Chung and Pennebaker, 2007), we may also expect that one person using first-person singular pronouns may cause the other to use more second-person pronouns, so that both people are referring to the same person. This is especially important under the Interactive Alignment Model view (Pickering and Garrod, 2004), where conversants align their entire mental representations, which predicts both lexical and referential alignment behaviors will be observed. Discourse-strategic explanations for alignment also predict alignment at multiple levels (Doyle and Frank, 2016).

Since we have access to a high-quality corpus with meaningful outcome measures, we can investigate the relative importance of these two types of alignment. We will show that referential alignment is more predictive of employment outcomes than is lexical alignment, suggesting a need for alignment research to consider both levels rather than just the latter.

### 3 Data: Corporate Email Corpus

We use the complete corpus of email between full-time employees at a mid-sized technology company between 2009 to 2014 (Srivastava et al., forthcoming). To protect privacy and confidentiality, all identifying information was removed, and each email was summarized as only a count of word categories in its text. These categories are a subset of the Linguistic Information and Word Count system (Pennebaker et al., 2007). The categories were chosen because they are likely to be indicative of one’s standing/role within a group.<sup>1</sup>

<sup>1</sup>Six pronoun categories (first singular (*I*), first plural (*we*), second (*you*), third singular personal (*he*, *she*), third singular impersonal (*it*, *this*), and third plural (*they*)) and five

We divided email chains into message-reply pairs to investigate conditional alignment between a message and its reply. To limit these pairs to cases where the reply was likely related to the preceding message, we removed all emails with more than one sender or recipient (including CC/BCC), identical sender and recipient, or where the sender or recipient was an automatic notification system or any other mailbox that was not specific to a single employee. We also excluded emails with no body text or more than 500 words in the body text, and pairs with more than a week’s latency between message and reply. Finally, because our analyses involve enculturation dynamics over the first six months of employment, we excluded replies sent by an employee whose overall tenure was less than six months. This resulted in a collection of 407,779 message-reply pairs, with 485 distinct replying employees. We combined this with monthly updates of employees joining and leaving the company and whether they left voluntarily or involuntarily. Of the 485, 66 left voluntarily, 90 left involuntarily, and 329 remained employed at the end of the observation period.

### 4 Model: An Extended WHAM Framework

To assess alignment, we use the Word-Based Hierarchical Alignment Model (WHAM) framework (Doyle and Frank, 2016). The core principle of WHAM is that alignment is a change, usually an increase, in the frequency of using a word category in a reply when the word category was used in the preceding message. For instance, a reply to the message *What will we discuss at the meeting?*, is likely to have more instances of future tense than a reply to the message *What did we discuss at the meeting?* Under this definition, alignment is the log-odds shift from the *baseline* reply frequency, the frequency of the word in a reply when the preceding message did **not** contain the word.

WHAM is a hierarchical generative modeling framework, so it uses information from related observations (e.g., multiple repliers with similar demographics) to improve its robustness on sparse data (Doyle et al., 2016). There are two key parameters, shown in Figure 2:  $\eta^{base}$ , the log-odds of a given word category  $c$  when the preceding message did not contain  $c$ , and  $\eta^{align}$ , the increase in

time/certainty categories (past tense, present tense, future tense, certainty, and tentativity).



the log-odds of  $c$  when the preceding message did contain  $c$ .

**A dynamic extension** To understand enculturation, we need to track changes in both the alignment and baseline over time. We add a month-by-month change term to WHAM, yielding a piecewise linear model of these factors over the course of an employee’s tenure. Each employee’s tenure is broken into two or three segments: their first six months after being hired, their last six months before leaving (if they leave), and the rest of their tenure.<sup>2</sup> The linear segments for their alignment are fit as an intercept term  $\eta^{align}$ , based at their first month (for the initial period) or their last month (for the final period), and per-month slopes  $\alpha$ . Baseline segments are fit similarly, with parameters  $\eta^{base}$  and  $\beta$ .<sup>3</sup> To visualize the alignment behaviors and the parameter values, we create “sawhorse” plots, with an example in Figure 1.

In our present work, we are focused on changes in cultural fit during the transitions into or out of the group, so we collapse observations outside the first/last six months into a stable point estimate, constraining their slopes to be zero. This simplification also circumvents the issue of different employees having different middle-period lengths.<sup>4</sup>

**Model structure** The graphical model for our instantiation of WHAM is shown in Figure 2. For each word category  $c$ , WHAM’s generative model represents each reply as a series of token-by-token independent draws from a binomial distribution. The binomial probability  $\mu$  is dependent on whether the preceding message did ( $\mu^{align}$ ) or did not ( $\mu^{base}$ ) contain a word from category  $c$ , and the inferred alignment value is the difference between these probabilities in log-odds space ( $\eta^{align}$ ).

<sup>2</sup>Within each segment, the employee’s alignment model is similar to that of Yurovsky et al. (2016), who introduced a constant by-month slope parameter to model changes in parent-child alignment during early linguistic development.

<sup>3</sup>Informal investigations into the change in baseline usage over time showed roughly linear changes over the first/last six months, but our linearity assumption may mask interesting variation in the enculturation trajectories that may reward further investigation.

<sup>4</sup>As shown in Figure 1, the pieces do not need to define a continuous function. Alignment behaviors continue to change in the middle of an employee’s tenure (Srivastava et al., forthcoming), so alignment six months in to the job is unlikely to be equal to alignment six months from leaving, or the average alignment over the middle tenure.

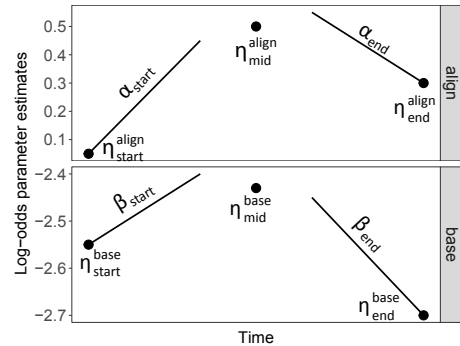


Figure 1: Sample sawhorse plot with key variables labelled. The  $\eta$  point parameters (first month, last month, and middle average) and  $\alpha$  (or  $\beta$ ) by-month slope (start, end) parameters are estimated by WHAM for each word category and employee group.

The specific values of these variables depend on three hierarchical features: the word category  $c$ , the group  $g$  that a given employee falls into, and the time period  $t$  (a piece of the piece-wise linear function: beginning, middle, or end). Note that the hierarchical ordering is different for the  $\eta$  chains and the  $\alpha/\beta$  chains;  $c$  is above  $g$  and  $t$  for the  $\eta$  chains, but below them for the  $\alpha/\beta$  chains. This is because we expect the static ( $\eta$ ) values for a given word category to be relatively consistent across different groups and at different times, but we expect the values to be independent across the different word categories. Conversely, we expect that the enculturation trajectories across word categories ( $\alpha/\beta$ ) will be similar, while the trajectories may vary substantially across different groups and different times. Lastly, the month  $m$  in which a reply is written (measured from the start of the time period  $t$ ) has a linear effect on the  $\eta$  value, as described below.

To estimate alignment, we first divide the replies up by group, time period, and calendar month. We separate the replies into two sets based on whether the preceding message contained the category  $c$  (the “alignment” set) or not (the “baseline” set). All replies within a set are then aggregated in a single bag-of-words representation, with category token counts  $C_{c,g,t,m}^{align}$  and  $C_{c,g,t,m}^{base}$ , and total token counts  $N_{c,g,t,m}^{base}$  and  $N_{c,g,t,m}^{align}$  comprising the observed variables on the far right of the model. Moving from right to left, these counts are assumed to come from binomial draws with probability  $\mu_{c,g,t,m}^{align}$  or  $\mu_{c,g,t,m}^{base}$ . The  $\mu$  values are then

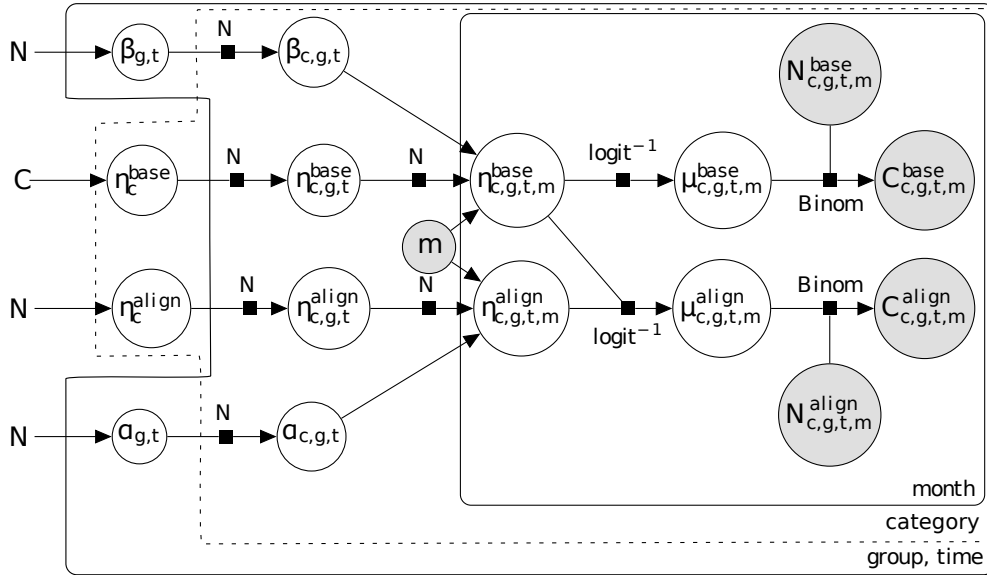


Figure 2: The Word-Based Hierarchical Alignment Model (WHAM). Hierarchical chains of normal distributions capture relationships between word categories, individuals, outcome groups, and time, and generate linear predictors  $\eta$ , which are converted into probabilities  $\mu$  for binomial draws of the words in replies.

in turn generated from  $\eta$  values in log-odds space by an inverse-logit transform, similar to linear predictors in logistic regression.

The  $\eta^{base}$  variables are representations of the baseline frequency of a marker in log-odds space, and  $\mu^{base}$  is simply a conversion of  $\eta^{base}$  to probability space, the equivalent of an intercept term in a logistic regression.  $\eta^{align}$  is an additive value, with  $\mu^{align} = \text{logit}^{-1}(\eta^{base} + \eta^{align})$ , the equivalent of a binary feature coefficient in a logistic regression. The specific month's  $\eta$  variables are calculated as a linear function:  $\eta_{c,g,t,m}^{align} = \eta_{c,g,t}^{align} + m\alpha_{c,g,t}$ , and similarly with  $\beta$  for the baseline.

The remainder of the model is a hierarchy of normal distributions that integrate social structure into the analysis. In the present work, we have three levels in the hierarchy: category, group, and time period. In Analysis 1, employees are grouped by their employment outcome (stay, leave voluntarily, leave involuntarily); in Analyses 2 & 3, where we predict the employment outcomes, each group is a single employee. The normal distributions that connect these levels have identical standard deviations  $\sigma^2 = .25$ .<sup>5</sup> The hierarchies are headed by a normal distribution centered at

<sup>5</sup>The deviation is not a theoretically motivated choice, and was chosen as a good empirical balance between reasonable parameter convergence (improved by smaller  $\sigma^2$ ) and good model log-probability (improved by larger  $\sigma^2$ ).

0, except for the  $\eta^{base}$  hierarchy, which has a  $Cauchy(0, 2.5)$  distribution.<sup>6</sup>

Message and reply length can affect alignment estimates; the WHAM model was developed in part to reduce this effect. As different employees had different email length distributions, we further accounted for length by dividing all replies into five quintile length bins, and treated each bin as separate observations for each employee. This design choice adds an additional control factor, but results were qualitatively similar without it. All of our analyses are based on parameter estimates from RStan fits of WHAM with 500 iterations over four chains.

While previous research on cultural fit has emphasized either its internalization (O'Reilly et al., 1991) or self-regulation (Goldberg et al., 2016) components, our extension to the WHAM framework helps disentangle them by estimating them as separate baseline and alignment trajectories. For example, we can distinguish between an archetypal individual who initially aligns to her colleagues and then internalizes this style of com-

<sup>6</sup>As  $\eta^{base}$  is the log-odds of each word in a reply being a part of the category  $c$ , it is expected to be substantially negative. For example, second person pronouns (*you*), are around 2% of the words in replies, approximately  $-4$  in log-odds space. We follow Gelman et al. (2008)'s recommendation of the Cauchy prior as appropriate for parameter estimation in logistic regression.

munication such that her baseline use also shifts and another archetypal person who aligns to her colleagues but does not change her baseline usage. The former exhibits high correspondence between internalization and self-regulation, whereas the latter demonstrates an ability to decouple them.

## 5 Analyses

We perform three analyses on this data. First, we examine the qualitative behaviors of pronoun alignment and how they map onto employee outcomes in the data. Second, we show that these qualitative differences in early enculturation are meaningful, with alignment behaviors predicting employment outcome above chance. Lastly, we consider lexical versus referential levels of alignment and show that predictions are improved under the referential formulation, suggesting that alignment is not limited to low-level word-repetition effects.

### 5.1 Analysis 1: Dynamic Qualitative Changes

We begin with descriptive analyses of the behavior of pronouns, which are likely to reflect incorporation into the company. In particular, we look at first-person singular (*I*), first-person plural (*we*), and second-person pronouns (*you*). We expect that increases in *we* usage will occur as the employee is integrated into the group, while *I* and *you* usage will decrease, and want to understand whether these changes manifest on baseline usage (i.e., internalization), alignment (i.e., self-regulation), or both.

**Design** We divided each employee’s emails by calendar month, and separated them into the employee’s first six months, their last six months (if an employee left the company within the observation period), and the middle of their tenure. Employees with fewer than twelve months at the company were excluded from this analysis, so that their first and last months did not overlap. We fit two WHAM models in this analysis. The first aggregated all employees, regardless of employment outcome, to minimize noise; the second separated them by outcome to analyze cultural fit differences.

**Outcome-aggregated model** We start with the aggregated behavior of all employees, shown in Figure 3. For baselines, we see decreased use of *I* and *you* over the first six months, with *we* usage

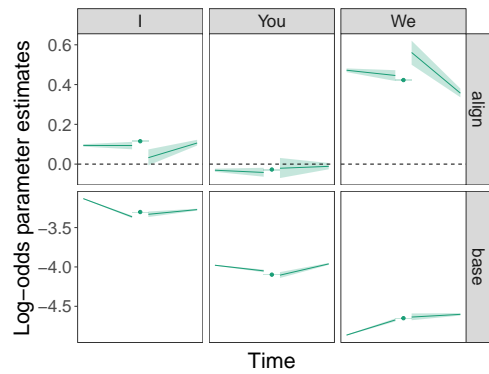


Figure 3: Sawhorse plots showing the dynamics of pronoun alignment behavior across employees. Vertical axis shows log-odds for baseline and alignment. Top row shows estimated alignment, highest for *we* and smallest for *you*. Bottom row shows baseline dynamics, with employees shifting toward the average usage as they enculturate. The shaded region is one standard deviation over parameter samples.

increasing over the same period, confirming the expected result that incorporating into the group is accompanied by more inclusive pronoun usage. Despite the baseline changes, alignment is fairly stable through the first six months. Alignment on first-person singular and second-person pronouns is lower than first-person plural pronouns, likely due to the fact that *I* or *you* have different referents when used by the two conversants, while both conversants could use *we* to refer to the same group. We will consider this referential alignment in more detail in Analysis 3. Since employees with different outcomes have much different experiences over their last six months, we will not discuss them in aggregate, aside from noting the sharp decline in *we* alignment near the end of the employees’ tenures.

**Outcome-separated model** Figure 4 shows outcome-specific trajectories, with green lines showing involuntary leavers (i.e., those who are fired or downsized), blue showing voluntary leavers, and orange showing employees who remained at the company through the final month of the data. The use of *I* and *you* is similar to the aggregates in Figure 3, regardless of group. The last six months of *I* usage show an interesting difference, where involuntary leavers align more on *I* but retain a stable baseline while voluntary leavers retain a stable alignment but increase *I* overall,

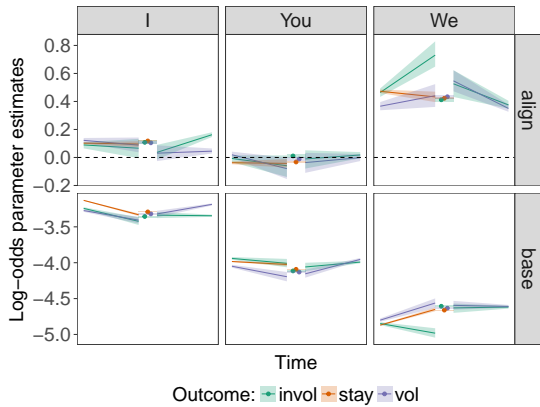


Figure 4: Sawhorse plots split by employment outcome. Mid-tenure points are jittered for improved readability.

which is consistent with group separation.

The most compelling result we see here, though, is the changes in *we* usage by different groups of employees. Employees who eventually leave the company involuntarily show signs of more self-regulation than internalization over the first six months, increasing their alignment while decreasing their baseline use (though they return to more similar levels as other employees later in their tenure). Employees who stay at the company, as well as those who later leave voluntarily, show signs of internalization, increasing their baseline usage to the company average, as well as adapting their alignment levels to the mean. This finding suggests that how quickly the employees internalize culturally-standard language use predicts their eventual employment outcome, even if they eventually end up near the average.

## 5.2 Analysis 2: Predicting Outcomes

This analysis tests the hypothesis that there are meaningful differences in employees’ initial enculturation, captured by alignment behaviors. We examine the first six months of communications and attempt to predict whether the employee will leave the company. We find that, even with a simple classifier, alignment behaviors are predictive of employment outcome.

**Design** We fit the WHAM model to only the first six months of email correspondence for all employees who had at least six months of email. The model estimated the initial level of baseline use ( $\eta_{base}$ ) and alignment ( $\eta_{align}$ ) for each employee, as well as the slope ( $\alpha, \beta$ ) for baseline and align-

ment over those first six months, over all 11 word categories mentioned in Section 3.

We then created logistic regression classifiers, using the parameter estimates to predict whether an employee would leave the company. We fit separate classifiers for leaving voluntarily or involuntarily. Our results show that early alignment behaviors are better at identifying employees who will leave involuntarily than voluntarily, consistent with (Srivastava et al., forthcoming)’s findings that voluntary leavers are similar to stayers until late in their tenure. We fit separate classifiers using the alignment parameters and the baseline parameters to investigate their relative informativity.

For each model, we report the area under the curve (AUC). This value is estimated from the receiver operating characteristic (ROC) curve, which plots the true positive rate against the false positive rate over different classification thresholds. An AUC of 0.5 represents chance performance. We use balanced, stratified cross-validation to reduce AUC misestimation due to unbalanced outcome frequencies and high noise (Parker et al., 2007).

**Results** The left column of Figure 5 shows the results over 10 runs of 10-fold balanced logistic classifiers with stratified cross-validation. The alignment-based classifiers are both above chance at predicting that an employee will leave the company, whether involuntarily or voluntarily. The baseline-based classifiers are above chance for involuntary leavers but are non-predictive for voluntary leavers. This finding is consistent with the idea that voluntary leavers resemble stayers (who form the bulk of the employees) until late in their tenure when their cultural fit declines.

We fit a model using both alignment and baseline parameters, but this model yielded an AUC value below the alignment-only classifier. This suggests that where alignment and baseline behaviors are both predictive, they do not provide substantially different predictive power and lead to overfitting. A more sophisticated classifier may overcome these challenges; our goal here was not to achieve maximal classification performance but to test whether alignment provided any useful information about employment outcomes.

## 5.3 Analysis 3: Types of Alignment

Our final analysis investigates the nature of linguistic alignment: specifically, whether there is an



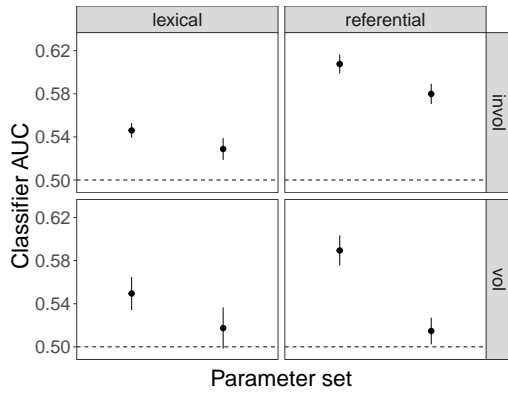


Figure 5: AUC values for 10 runs of 10-fold cross-validated logistic classifiers. Both lexical (left column) and referential (right column) alignment parameters lead to above chance classifier performance, but referential alignment outperforms lexical alignment at predicting both voluntary and involuntary departures.

effect of referential alignment beyond that of the more commonly used lexical alignment.

Testing this hypothesis requires a small change to the alignment calculations. Lexical alignment is based on the conditional probability of the replier using a word category  $c$  given that the preceding message used that same category  $c$ . For referential alignment, we examine the conditional probability of the replier using a word category  $c_j$  given that the preceding message used the category  $c_i$ , where  $c_i$  and  $c_j$  are likely to be referentially linked. We also consider cases where  $c_i$  is likely to transition to  $c_j$  throughout the course of the conversation, such as present tense verbs turning into past tense as the event being described recedes into the past. The pairs of categories that are likely to be referentially or transitionally linked are: (*you*, *I*); (*we*, *I*); (*you*, *we*); (past, present); (present, future); and (certainty, tentativity). We include both directions of these pairs, so this provides approximately the same number of predictor variables for both situations to maximize comparability (12 for the referential alignments, 11 for the lexical). This modification does not change the structure of the WHAM model, but rather changes its  $C$  and  $N$  counts by reclassifying replies between the baseline or alignment pathways.

**Results** Figure 5 plots the differences in predictive model performance using lexical versus referential alignment parameters. We find that the semantic parameters provide more accurate clas-

sification than the lexical both for voluntarily and involuntarily-leaving employees. This suggests that while previous work looking at lexical alignment successfully captures social structure, referential alignment may reflect a deeper and more accurate representation of the social structure. It is unclear if this behavior holds in less formal situations or with weaker organizational structure and shared goals, but these results suggest that the traditional alignment approach of only measuring lexical alignment should be augmented with referential alignment measures for a more complete analysis.

## 6 Conclusions

This paper described an effort to use directed linguistic alignment as a measure of cultural fit within an organization. We adapted a hierarchical alignment model from previous work to estimate fit within corporate email communications, focusing on changes in language during employees’ entry to and exit from the company. Our results showed substantial changes in the use of pronouns, with pronoun patterns varying by employees’ outcomes within the company. The use of the first-person plural “we” during an employee’s first six months is particularly instructive. Whereas stayers exhibited increased baseline use, indicating internalization, those eventually departing involuntarily were on the one hand decreasingly likely to introduce “we” into conversation, but increasingly responsive to interlocutors’ use of the pronoun. While not internalizing a shared identity with their peers, involuntarily departed employees were overly self-regulating in response to its invocation by others.

Quantitatively, rates of usage and alignment in the first six months of employment carried information about whether employees left involuntarily, pointing towards fit within the company culture early on as an indicator of eventual employment outcomes. Finally, we saw ways in which the application of alignment to cultural fit might help to refine ideas about alignment itself: preliminary analysis suggested that referential, rather than lexical, alignment was more predictive of employment outcomes. More broadly, these results suggest ways that quantitative methods can be used to make precise application of concepts like “cultural fit” at scale.

## References

- Richard Y. Bourhis and Howard Giles. 1977. The language of intergroup distinctiveness. In H Giles, editor, *Language, Ethnicity, and Intergroup Relations*, Academic Press, London, pages 119–135.
- David B. Buller and R. Kelly Aune. 1988. The effects of vocalics and nonverbal sensitivity on compliance: A speech accommodation theory explanation. *Human Communication Research* 14:301–32.
- Cindy Chung and James W. Pennebaker. 2007. The psychological functions of function words. In K Fiedler, editor, *Social communication*, Psychology Press, New York, chapter 12, pages 343–359.
- 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*, page 699. <https://doi.org/10.1145/2187836.2187931>.
- Cristian Danescu-Niculescu-Mizil, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. No country for old members: user lifecycle and linguistic change in online communities. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 307–318.
- Gabriel Doyle and Michael C. Frank. 2016. Investigating the sources of linguistic alignment in conversation. In *Proceedings of ACL*.
- Gabriel Doyle, Dan Yurovsky, and Michael C. Frank. 2016. A robust framework for estimating linguistic alignment in Twitter conversations. In *Proceedings of WWW*.
- R. E. Feldman. 1968. Response to compatriots and foreigners who seek assistance. *Journal of Personality and Social Psychology* 10:202–14.
- Riccardo Fusaroli, Bahador Bahrami, Karsten Olsen, Andreas Roepstorff, Geraint Rees, Chris Frith, and Kristian Tylén. 2012. Coming to Terms: Quantifying the Benefits of Linguistic Coordination. *Psychological Science* 23(8):931–939. <https://doi.org/10.1177/0956797612436816>.
- Andrew Gelman, Aleks Jakulin, Maria Grazia Pittau, and Yu-Sung Su. 2008. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*.
- Howard Giles, Nikolas Coupland, and Justine Coupland. 1991. Accommodation theory: Communication, context, and consequences. In Howard Giles, Justine Coupland, and Nikolas Coupland, editors, *Contexts of accommodation: Developments in applied sociolinguistics*, Cambridge University Press, Cambridge.
- Augusto Gnisci. 2005. Sequential strategies of accommodation: A new method in courtroom. *British Journal of Social Psychology* 44(4):621–643.
- Amir Goldberg, Sameer B. Srivastava, V. Govind Manian, and Christopher Potts. 2016. Fitting in or standing out? The tradeoffs of structural and cultural embeddedness. *American Sociological Review*.
- Molly E. Ireland, Richard B. Slatcher, Paul W. Eastwick, Lauren E. Scissors, Eli J. Finkel, and James W. Pennebaker. 2011. Language style matching predicts relationship initiation and stability. *Psychological Science* 22:39–44. <https://doi.org/10.1177/0956797610392928>.
- Kate G. Niederhoffer and James W. Pennebaker. 2002. Linguistic style matching in social interaction. *Journal of Language and Social Psychology* 21(4):337–360. <http://jls.sagepub.com/content/21/4/337.short>.
- Charles A. O’Reilly, Jennifer Chatman, and David F. Caldwell. 1991. People and organizational culture: a profile comparison approach to assessing person-organization fit. *Academy of Management Journal* 34(3):487–516.
- Brian J Parker, Simon Günter, and Justin Bedo. 2007. Stratification bias in low signal microarray studies. *BMC bioinformatics* 8(1):1.
- James W. Pennebaker, Cindy K. Chung, Molly Ireland, Amy Gonzalez, and Roger J. Booth. 2007. The development and psychometric properties of liwc2007. Technical report, LIWC.net.
- Martin J. Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences* 27(2):169–190. <https://doi.org/10.1017/S0140525X04000056>.
- Sameer B. Srivastava, Amir Goldberg, V. Govind Manian, and Christopher Potts. forthcoming. Enculturation trajectories: Language, cultural adaptation, and individual outcomes in organizations. *Management Science*.
- Harry C. Triandis. 1960. Cognitive similarity and communication in a dyad. *Human Relations* 13:175–183.
- Dan Yurovsky, Gabriel Doyle, and Michael C. Frank. 2016. Linguistic input is tuned to children’s developmental level. In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*.