Investigating the Components of Linguistic Alignment

Anonymous ACL submission

Abstract

In conversation, speakers tend to "accommodate" or "align" to their partners, changing the style and substance of their communications to be more similar to their partners. We focus here on "linguistic alignment," changes in word choice based on others' choices. Although linguistic alignment is observed across many different contexts and its degree correlates with important social factors such as power and likability, its origins are still uncertain. We build on a recent probabilistic model of alignment, using it to separate out the influence of words and word categories. We then compare alignment across contexts (telephone conversations vs. microblog replies). Our results suggest that alignment is consistently higher for categories with specific discourse functions. Furthermore, we find that different discourse acts also show different degrees of alignment. Rather than stemming from non-communicative priming processes, alignment arises from and interacts with the ongoing discourse.

1 Introduction

In conversation, people tend to adapt to one another across a broad range of behaviors. This adaptation behavior is collectively known as "communication accommodation" (Giles et al., 1991). Linguistic alignment, the use of similar words to a conversational partner, is one prominent form of accommodation. Alignment is found robustly across many settings, including in-person, computer-mediated, and web-based conversation (Danescu-Niculescu-Mizil et al., 2012; Giles et al., 1979; Niederhoffer and Pennebaker, 2002). In addition, the strength of alignment to conversational partners varies with important sociological factors, such as the power of the partners, their social network centrality, and their likability. Linguistic alignment might even be used to infer these factors in situations where they are unobserved.

Although linguistic alignment appears to reflect important social dynamics, the mechanisms underlying alignment are still not well-understood. We focus here on whether alignment is supported by lower-level, relatively automatic priming mechanisms, or higher-level, strategic discourse processes. Arguing for a priming-based mechanism, the Interactive Alignment Model proposes that conversational partners prime each other, causing alignment via the primed reuse of structures ranging from individual lexical items to syntactic abstractions (Pickering and Garrod, 2004). In contrast, Accommodation Theory emphasizes the relatively more communicative and strategic nature of alignment (Giles et al., 1991).

Although they are not quantitative theories, these accounts nevertheless make a number of testable predictions. First, if alignment is driven by priming, effects should be relatively constant across different words and word categories. In contrast, communicative accounts should predict that alignment could differ, perhaps being more focused on words, categories, and structures that serve particular conversational or discourse functions. Second, if alignment is driven by priming, it should be relatively consistent across different aspects of a discourse. In contrast, from a strategic or communicative perspective, alignment – in which preceding words and concepts are reused – must be balanced against a need to move the conversation forward by introducing new words and concepts. Thus, on a communicative account, alignment should be modulated by the speaker's

discourse act, reflecting whether the balance of the concern is convergence on a current focus or the conveyal of new information.

Our goal in the current work is to test these predictions. We make use of a recent probabilistic model of linguistic alignment, modifying it to operate robustly over corpora with highly varying distributional structures and to consider both lexical and category-based alignment. We use two corpora of spontaneous conversations, the Switchboard Corpus and a corpus of Twitter conversations, to perform two experiments. First, in both datasets we measure alignment across different levels of representation and across different categories. Alignment is consistently higher for words and categories with particular discourse functions, but it varies substantially by representation level. Second, we make use of discourse annotations in Switchboard to measure alignment across different discourse acts, finding that alignment is highly dependent on what discourse actions are being undertaken. Taken together, while these findings reaffirm a large lexical component for alignment, they also suggest the operation of discourse-level, strategic processes.

2 Previous Work

2.1 Why does alignment matter?

Linguistic alignment, like other kinds of accommodation, can be a critical part of achieving social goals. Performance in cooperative decision-making tasks is positively related to the participants' linguistic convergence (Fusaroli et al., 2012; Kacewicz et al., 2013). Romantically, match-making in speed dating and stability in established relationships have both been linked to increased alignment (Ireland et al., 2011). Alignment can also improve perceived persuasiveness, encouraging listeners to follow good health practices (Kline and Ceropski, 1984) or to leave larger tips (van Baaren et al., 2003).

Alignment is also important as an indicator of implicit sociological variables. Less powerful conversants generally accommodate to more to powerful conversants. Prominent examples include interviews and jury trials (Willemyns et al., 1997; Gnisci, 2005; Danescu-Niculescu-Mizil et al., 2012). A similar effect is found for network structure: speakers align more to more network-central speakers (Noble and Fernández, 2015). Additionally, factors such as gender, likability, respect, and attraction all interact with the magnitude of accommodation (Bilous and Krauss, 1988; Natale, 1975). Such differences in accommodation can also be indicative of changes to the power dynamic: In U.S. Supreme Court transcripts, (Guo et al., 2015) showed that depending on the accommodation dimension, justices – who are more powerful by any intuitive assessment – may nevertheless accommodate more to lawyers, perhaps because the lawyers have the local power to answer justices' questions.

2.2 What's the nature of alignment?

While linguistic alignment has been studied in multiple ways, the most prominent strand has focused on the level of word categories, looking at how interlocutors change their frequency of using, for instance, pronouns or quantitative words (Danescu-Niculescu-Mizil et al., 2012; Ireland et al., 2011). This line of work has not directly investigated whether alignments go beyond the level of individual lexical items and extend to syntactic structures. When studies have attempted to test for syntactic alignment, results have been equivocal. While some have found support for syntactic priming (Gries, 2005; Dubey et al., 2005), others have found negative or null alignment (Healey et al., 2014; Reitter et al., 2010).

In addition, as noted before, the psychological mechanism or mechanisms that lead to alignment is still uncertain. The Interactive Alignment Model (Pickering and Garrod, 2004) makes the theoretical claim that alignment should be similar across all structural levels as conversants increasingly share their representation. However, (Healey et al., 2014) find across two corpora that speakers syntactically *diverge* from their interlocutors once lexical alignment is accounted for.

Furthermore, positive alignment is treated as an inherently good thing, but there is clearly a limit to its goodness, as alignment is inherently backward-looking, while the general goal of a conversation is to exchange information that is not already known by both parties, and inherently forward-looking goal. There are suggestive results that alignment based on function words, which can stay constant even as the topic changes, is more appropriate than alignment based on all words for predicting performance in a decision-making task (Fusaroli et al., 2012). In addition, some recent work finding positive accommodation has limited

itself to "non-topical" word categories (Danescu-Niculescu-Mizil et al., 2011; Doyle et al., 2016).

In sum, because most work on alignment has been done either on categories of words or aggregating across the lexicon, we do not have a good sense of whether there are systematic differences between alignment at different levels. Hence, the evidence is equivocal on the precise nature and scope of alignment behaviors.

3 Measures of alignment

A further complication is that there is no one standard measure of alignment. The metrics used in previous work fall into two basic categories: distributional and conditional. Distributional methods such as Linguistic Style Matching (LSM) (Niederhoffer and Pennebaker, 2002; Ireland et al., 2011) or the Zelig Quotient (Jones et al., 2014) calculate the similarity between the conversation participants over their frequencies of word or word category use. In contrast, conditional metrics, such as Local Linguistic Alignment (LLA) (Fusaroli et al., 2012; Wang et al., 2014) and the metric used by (Danescu-Niculescu-Mizil et al., 2011), look at how a message conditions its reply, with convergence indicated by elevated word use in the reply when that word was in the preceding message.

While distributional methods have been popular, a major weakness of such methods is that they do not necessarily show true alignment, only similarity. A high level of distributional similarity does not imply that two conversational partners have aligned to one another, because they might instead have been similar to begin with. In contrast, conditional measures allow for stronger inferences about the temporal sequence of alignment (even though they cannot guarantee any causal interpretation). The work reported here extends a recent conditional metric, the Hierarchical Alignment Model, or HAM (Doyle et al., 2016).

By-message conditional methods Existing conditional methods have typically taken a binary view of utterances (Danescu-Niculescu-Mizil et al., 2012; Doyle et al., 2016). Consider the following example of conditional alignment on pronouns. Bob aligns to Alice if his replies are more likely to contain a pronoun when in response to a message from Alice that contains a pronoun. An example of positive conditional alignment is shown below:

	Bob's reply	
Alice's message	has pronoun	no pronoun
has pronoun	8	2
no pronoun	5	5

Here, Alice sends 10 messages that contain at least one pronoun, and 8 of Bob's replies contain at least one pronoun. But Alice also sends 10 messages that don't contain any pronouns, and only 5 of Bob's replies to these contain pronouns. This increased likelihood of a pronoun-containing reply to a pronoun-containing message is the conditional alignment.

Different models quantify this conditional alignment slightly differently. (Danescu-Niculescu-Mizil et al., 2011) proposed a subtractive conditional probability model, where alignment is the difference between the likelihood of a pronoun-containing reply to a pronoun-containing message and the probability of a pronoun-containing reply to any message:

$$align_{SCP} = p(B|A) - p(B) \quad (1)$$

Doyle, Yurovsky, and Frank (2016) showed that this measure can be affected by the overall frequency of the category being aligned on, though. To correct this issue, they proposed a Hierarchical Alignment Model (HAM), which defines alignment as a linear effect on the log-odds of a reply containing the relevant marker (e.g., a pronoun), similar to a linear predictor in a logistic regression.[1]

$$align_{HAM} \approx logit^{-1}(p(B|A)) - logit^{-1}(p(B|\neg A)) \quad (2)$$

Both of these binarized conditional methods, though, depend on the assumption that all messages have similar, and small, numbers of words. The probability that a message contains at least one of any marker of interest is of course dependent on the message's length, so if messages vary substantially in their length, these alignment values can be at least noisy, if not biased. They are also not robust as messages increase in length, since the likelihood that a message contains any marker approaches 1 as message length increases.

[1] Because the HAM estimated this quantity via Bayesian inference, the inferred alignment value depends on the prior and number of messages observed, so unlike the other measures, this equality is only approximate.

By-word conditional methods A solution to this problem is simply to shift from binarized data to count data. Instead of modeling whether or not a reply contains a marker of interest, we can model that marker's frequency within replies to messages that do or do not contain the marker. Some existing measures use the proportion of the preceding message that appears in its reply to estimate alignment, notably Local Linguistic Alignment (LLA) (Fusaroli et al., 2012; Wang et al., 2014) and the lexical similarity (LS) measure of (Healey et al., 2014). LLA is defined as the number of word tokens (w_i) that appear in both the message (M_a) and the reply (M_b), divided by the product of the total number of word tokens in the message and reply:

$$align_{LLA} = \frac{\sum_{w_i \in M_b} \delta(w_i \in M_a)}{length(M_a) length(M_b)} \quad (3)$$

These measures have an aspect of conditionality, as they only count words that appear in both the message and the reply, but they fail to control for the baseline frequency of the initial marker, and hence may be biased in measurements across words or categories of different frequencies (Doyle et al., 2016). They also have inherent length dependence, as the maximum alignment estimate is only possible when the reply is shorter than the message.[2]

We need, therefore, an alignment measure that has the benefits of the existing by-message conditional models (HAM outperformed SCP, LLA, and LSM in simulations in (Doyle et al., 2016)) while gaining the length-robustness of a by-word conditional method. A simple change to the HAM framework satisfies this goal.

4 The Word-Based Hierarchical Alignment Model (WHAM)

The HAM framework conceptualizes alignment as the increase in the likelihood that a reply will contain a "marker"—a word or word category—given that the preceding message contained it, compared to when the preceding message did not contain it. This introduces a substantial simplification: it treats each message as a binary variable, either containing or not containing the word of interest. By-message alignment has successfully found alignment effects in previous work (Danescu-Niculescu-Mizil et al., 2011), and may be appropriate for data such as the Twitter dataset used in that paper. However, in looking at more natural dialogues, such as telephone conversations, length may vary substantially between messages, obscuring the alignment effect.

We propose the Word-Based Hierarchical Alignment Model (WHAM) to address this. Like HAM, the WHAM framework assumes that word use in replies is shaped by whether the preceding message contained the marker of interest. But the WHAM framework looks at the marker token frequencies within the replies, so that a 40-word reply with two instances of the marker is represented differently from a 3-word reply containing with one instance.

For each marker, WHAM treats each reply treated as a sample of token-by-token independent draws from a binomial distribution. The binomial probability is dependent on whether the preceding message did (μ^{align}) or did not (μ^{base}) contain the marker, and the inferred alignment value is the difference between these probabilities in log-odds space (η^{align}).

WHAM's graphical model is shown in Figure 1. For a set of message-reply pairs between a speaker-replier dyad (a, b), we first separate the replies into two sets based on whether the preceding message contained the marker m (the "alignment" set) or not (the "baseline" set). All replies within a set are then aggregated in a single bag-of-words representation, with marker token counts $C^{align}_{m,a,b}$ and $C^{base}_{m,a,b}$, and total token counts $N^{base}_{m,a,b}$ and $N^{base}_{m,a,b}$, 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 $\mu^{align}_{m,a,b}$ or $\mu^{base}_{m,a,b}$. The μ values are generated from η values in log-odds space by an inverse-logit transform, similar to linear predictors in logistic regression.

The η^{base} variables are representations of the baseline frequency of a marker in log-odds space, and μ^{base} is simply a conversion of η^{base} to probability space, the equivalent of an intercept term in a logistic regression. η^{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. Alignment is the change in log-odds of the replier using m above their baseline usage of the marker, in response to a message that uses m.

[2] Proof in Supplemental?

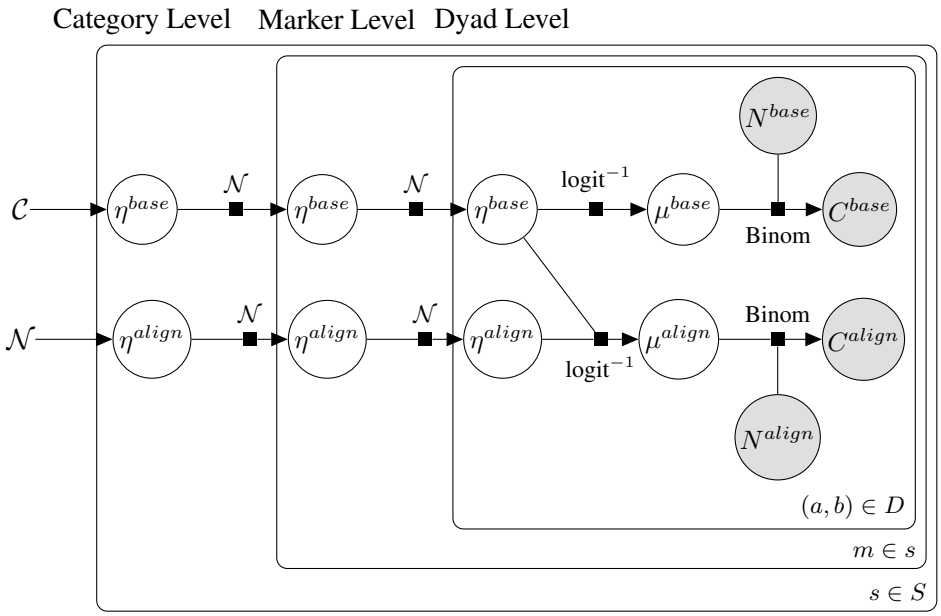

Figure 1: The Word-Based Hierarchical Alignment Model (WHAM). A chain of normal distributions generates a linear predictor η, which is converted into a probability μ for binomial draws of the words in each reply.

The remainder of the model is a hierarchy of normal distributions that bring category and social structure into the model, and as in HAM, this can vary depending on the problem the model is applied to. In the present work, we have three levels in the hierarchy: first, a category level, followed by a marker level[3], followed by a marker-dyad level. All of these normal distributions have identical standard deviations $\sigma^2 = .25$.[4] The baseline hierarchy is headed by a $Cauchy(0, 2.5)$ distribution, following (Gelman et al., 2008)'s suggestion, as fairly uninformative prior for baseline marker frequency. The alignment hierarchy is headed by a normal distribution centered at 0, biasing the model equally in favor of positive and negative alignments.

5 Experiments

Using the WHAM framework, we investigate the variation in alignment across different conversation contexts and across different discourse acts.

[3]In the lexical and category-not-word alignment models, these markers are words within a category. The category alignment model does not include this level, since all words in a category are treated identically.

[4]This value was chosen as a good balance between reasonable parameter convergence (improved by smaller σ^2) and good model log-probability (improved by larger σ^2).

Category	Examples	Size	Swbd Prob	Twit Prob
Article	*a, an, the*	2	.053	.047
Certainty	*always, never*	17	.014	.015
Conjunction	*but, and, though*	18	.077	.051
Discrepancy	*should, would*	21	.015	.019
Exclusive	*without, exclude*	77	.038	.028
Inclusive	*with, include*	57	.057	.028
Negation	*not, never*	12	.020	.023
Preposition	*to, in, by, from*	97	.097	.091
Pronoun	*it, you*	55	.17	.16
Quantifier	*few, many*	23	.028	.025
Tentative	*maybe, perhaps*	28	.033	.025

Table 1: Marker categories for linguistic alignment, with examples, number of distinct word lemmas, and token probability of in a reply in Switchboard and Twitter.

5.1 Data

We use two corpora in these experiments. The first is a collection of Twitter conversations collected by (Doyle and Frank, 2015) to examine information density in conversation. This corpus focuses on conversations within a set of 14 mostly distinct subcommunities on Twitter, and contains 63,673 conversation threads, covering 228,923 total tweets. We divide these conversations into message pairs, also called conversational turns, which are two consecutive tweets within a conversation thread. The second tweet is always in reply to the first (according to the Twitter API),

although this does not necessarily mean that the content of the reply is a response to the preceding tweet. Retweets (including explicit retweets and some common manual retweet methods) were removed automatically. This processing leaves us with 122,693 message pairs, spanning 2,815 users. The tweets were parsed into word tokens using the Twokenizer (Owoputi et al., 2013).

The second corpus is the SwDA version of the Switchboard corpus (**?**), as compiled by Chris Potts[5] . This is a collection of transcribed telephone conversations with each utterance labeled with the discourse act it is performing (e.g., statement of opinion, signal of non-understanding). It contains 221,616 total utterances in 1,155 conversations. We combine consecutive utterances by the same speaker without interruption from the listener into a single message and treat consecutive pairs of messages from different speakers as conversation turns. [XXX nnumber of conversatiuonal turns]

For both corpora, we use the Linguistic Inquiry and Word Count (LIWC) system to categorize words (Pennebaker et al., 2007). We use a set of 11 categories that have shown alignment effects in previous work (Danescu-Niculescu-Mizil et al., 2011). These can be loosely grouped into a set of five syntactic categories (articles, conjunctions, prepositions, pronouns, and quantifiers) and six conceptual categories (certainty, discrepancy, exclusion, inclusion, negation, and tentative). Categories and example elements are shown in Table **??**. The words within these categories were manually lemmatized.

5.2 Experiment 1

Our first experiment examines how alignment differs across the lexical and categorical levels.

5.2.1 Levels of alignment and their interaction

We want to examine the interaction of two levels of alignment. The first is lexical-level alignment, looking at how the presence of a word (lemma) in a message increases its likelihood of appearing in the reply. The second is category-level alignment, looking at how the presence of a member of a word category in a message increases the likelihood of words from that category appearing in the reply.

[5] http://compprag.christopherpotts.net/swda.html

	Reply		
Message	$\emptyset$	*he*	*she*
$\emptyset$	25	25	25
he	20	50	10
she	20	10	50

Table 2: A theoretical case where lexical alignment surpasses categorical alignment due to non-independence between the words.

These categories generally correspond to syntactic classes (pronouns, conjunctions) or broad conceptual classes (inclusive words, negative emotions) Work on by-word conditional alignment has tended to focus on the lexical level, while distributional and by-message conditional alignment work has focused on the categorical level.

Furthermore, we are interested in explicitly tracking the influence of lexical alignment on categorical alignment. It is possible that the category alignment effects in previous work are the result of lexical alignment on the individual words in the category, without any influence across words in the category. If categorical alignment is a real effect over and above lexical alignment, then the presence of a word in a message should not only increase the chance of seeing that word in the reply, but also other words in its category.

However, assessing the amount of alignment triggered across words in a category (which we call "category-not-word alignment") is not trivial, as there are a variety of interactions between lexical items within a category that can cause the lexical alignment to actually be less than the category alignment. Table 2 illustrates this with a theoretical distribution over the pronouns *he* and *she*; one use of the pronoun *he* makes another use more likely (*Did he like the movie? Yeah, he loved it.*) while also reducing the likelihood of *he*, since the topic of conversation is now a male, and vice versa for *she*. For both *he* and *she*, the lexical alignment is approximately $logit^{-1}(p(B|A) - p(B|\neg A)) = logit^{-1}(\frac{50}{80} - \frac{25}{75}) \approx 1.2$, but categorical alignment is approximately $logit^{-1}(\frac{120}{160} - \frac{50}{75}) \approx 0.4$.

Lexical alignment can differ from categorical alignment for a variety of reasons, so we consider three quantities: the lexical and categorical alignments, but also the "category-not-word" (CNW) alignment: the increased likelihood of seeing another member of the category other than the target word.

	Reply		
Message	$\emptyset$	*you*	*I*
$\emptyset$	25	25	25
you	20	10	50
I	20	50	10

Table 3: A theoretical case where lexical alignment is below categorical alignment due to non-independence between the words.

5.3 Category-not-word Alignment

To investigate CNW alignment, we look at a subset of the data: for each word w, exclude all messages that contain a word from that category that is not w. This limits the category alignment influence on the reply to the single word w. Then, instead of looking at how often w appears in the reply, we look at how often all other words in the category S appear in the reply. The model then infers the influence of w on the other words in the category independent of their lexical alignment.

This is implemented within the WHAM framework by changing the count variables $C^{\cdot}$ and $N^{\cdot}$. C^{align} is the number of tokens of $\{S - w\}$ in replies to messages containing w but not $\{S - w\}$. C^{base} is the number in replies to messages not containing any words in S. Similarly, N^{align} is the total token counts over replies containing w but not any other words in S, and N^{base} the total token counts over replies containing no words in S.

The key quantity in our analyses is the inferred marker-level variable η_S^{align}, the level of alignment on category S across all dyads. For lexical and CNW alignments, this is essentially a weighted average over the different lexical items in the category.

5.3.1 Results

We implemented WHAM in RStan (Carpenter, 2015), with code available at `http://github.com/langcog/alignment`. The model is fit with two chains of 200 iterations of the sampler (100 discarded as burn-in) for each dataset; judging from trace plots, this setting led to reliable convergence. We then extracted alignment estimates from each of the final 100 iterations of the model, and we report the 95% highest posterior density interval on the parameter values in these plots.

We find that there are substantial differences across contexts in the overall rate of alignment between the corpora (mean category alignment on

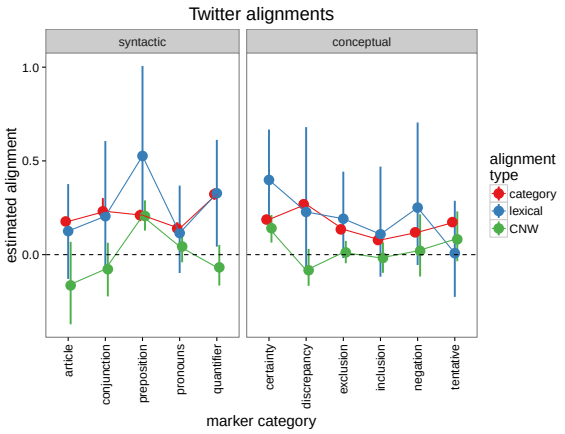

Figure 2: Categorical (red), lexical (blue), and CNW (green) alignments on the Twitter dataset. 95% HPD intervals from WHAM shown.

Twitter was .19, while its mean on Switchboard was $-.051$). This may reflect the nature of the two discourses. Replies on Twitter are composed while looking at the preceding message, encouraging the replier to take more account of the other tweeter's words, and a replier can draft and edit their reply to make it better fit the conversation. Messages on Switchboard, on the other hand, are evanescent, so a replier must compose a reply without looking back at the message, without editing, and in real-time. In Experiment 2, we will show that differences in discourse structure may also impact the differences in alignment.

Despite these differences, there is a crucial similarity: for both corpora, the strength of lexical alignment is significantly larger than the CNW alignment, according to a t-test over all categories (Twitter: $t(10) = .21, p < .001$; Swbd: $t(10) = .12, p = .003$), while lexical and categorical alignment are not significantly different. This suggests that alignment on these categories is substantially lexical, even on categories like pronouns where higher CNW alignments may have been expected.

The two corpora also differ in the correlations between the alignment measures. The Switchboard corpus, despite its overall tendency against category alignment, has significant positive correlations between category and lexical ($r = .71, p = .015$), and category and CNW alignment ($r = .94, p < .001$), as well as marginal positive alignment between lexical and CNW alignment ($r = .52, p = .10$). This suggests a consistency across words within a category.

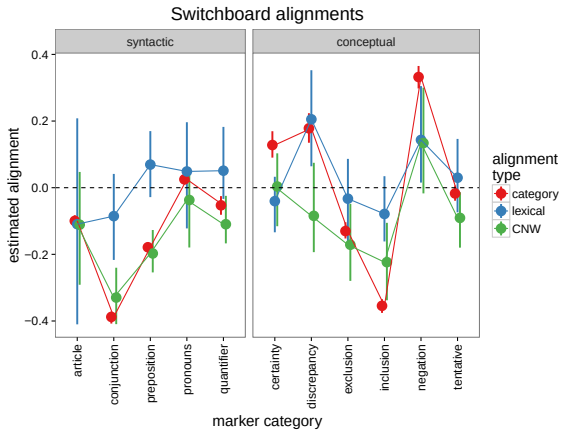

Figure 3: Categorical (red), lexical (blue), and CNW (green) alignments on the Switchboard dataset. 95% HPD intervals from WHAM shown.

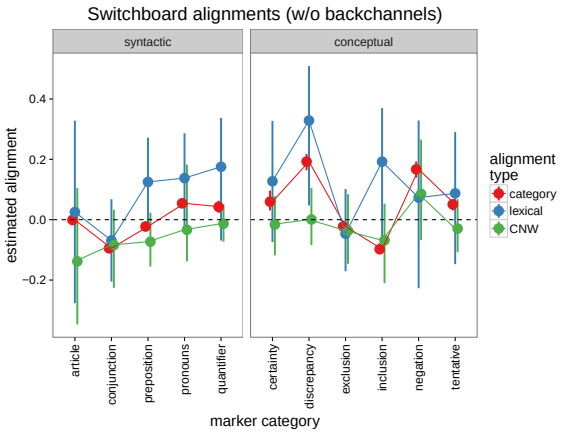

Figure 4: Categorical (red), lexical (blue), and CNW (green) alignments on the Switchboard dataset with backchannels removed. 95% HPD intervals from WHAM shown.

5.4 Experiment 2

The second experiment examines how alignment differs depending on discourse act. In particular, we focus on how the presence of backchannels in Switchboard may account for some of the differences between Twitter and Switchboard.

5.4.1 Discourse Acts

Different messages within a discourse serve different purposes, and this affects both their linguistic structure and their relationship with neighboring messages. A simple yes/no question is likely to receive a short, constrained reply, while a statement of an opinion is more likely to yield a longer reply. In addition, different types of messages can either introduce new information to the conversation (e.g., statements, questions, offers) or look back at existing information (e.g., acknowledgments, reformulations, yes/no answers). We hypothesize that alignment will be substantially different depending on the discourse act, as speakers' conversational goals vary.

We focus on a particular kind of discourse act, the backchannel (**?**). Backchannels are highly common in Switchboard, accounting for almost 20% of utterances, and include utterances such as single words signaling understanding or misunderstanding (*yeah, uh-huh*) and simple messages expressing mild emotional resonance (*It must have been tough*). Backchannels are a particularly interesting case because on some categories it is important to align (e.g., matching the tentative or negative tone of a speaker) while in other categories it may be very difficult (e.g., backchannels rarely contain quantifiers or prepositions).

In this experiment, we compare alignment in conversations containing backchannels with those whose backchannels have been removed. This is done by removing every utterance classified as a backchannel from the corpus and parsing the utterances into conversation turns as before.

5.4.2 Results

References

FR Bilous and RM Krauss. 1988. Dominance and accommodation in the conversational behaviours of same-and mixed-gender dyads. *Language & Communication*.

Bob Carpenter. 2015. Stan: A Probabilistic Programming Language. *Journal of Statistical Software*.

Cristian Danescu-Niculescu-Mizil, Michael Gamon, and Susan Dumais. 2011. Mark my words!: linguistic style accommodation in social media. In *Proceedings of the 20th international conference on World Wide Web - WWW '11*, page 745, New York, New York, USA. ACM Press.

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 - WWW '12*, page 699.

G Doyle and M C Frank. 2015. Audience size and contextual effects on information density in Twitter conversations. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*.

Gabriel Doyle, Dan Yurovsky, and Michael C. Frank. 2016. A robust framework for estimating linguistic alignment in Twitter conversations. In *WWW 2016*.

Amit Dubey, Patrick Sturt, and Frank Keller. 2005. Parallelism in coordination as an instance of syntactic priming: Evidence from corpus-based modeling. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 827–834. Association for Computational Linguistics.

R. Fusaroli, B. Bahrami, K. Olsen, A. Roepstorff, G. Rees, C. Frith, and K. Tylen. 2012. Coming to Terms: Quantifying the Benefits of Linguistic Coordination. *Psychological Science*, 23(8):931–939.

A Gelman, A Jakulin, MG Pittau, and YS Su. 2008. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*.

H Giles, K R Scherer, and D M Taylor. 1979. Speech markers in social interaction. In K R Scherer and H Giles, editors, *Social markers in speech*, pages 343–81. Cambridge University Press, Cambridge.

H Giles, N Coupland, and J Coupland. 1991. Accommodation theory: Communication, context, and consequences. In H Giles, J Coupland, and N Coupland, editors, *Contexts of accommodation: Developments in applied sociolinguistics*. Cambridge University Press, Cambridge.

A Gnisci. 2005. Sequential strategies of accommodation: A new method in courtroom. *British Journal of Social Psychology*, 44(4):621–643.

Stefan Th Gries. 2005. Syntactic priming: A corpus-based approach. *Journal of psycholinguistic research*, 34(4):365–399.

Fangjian Guo, Charles Blundell, Hanna Wallach, Katherine Heller, and U C L Gatsby Unit. 2015. The Bayesian Echo Chamber: Modeling Social Influence via Linguistic Accommodation. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, pages 315–323.

Patrick GT Healey, Matthew Purver, and Christine Howes. 2014. Divergence in dialogue. *PloS one*, 9(6):e98598.

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.

S. Jones, R. Cotterill, N. Dewdney, K. Muir, and A. Joinson. 2014. Finding Zelig in text: A measure for normalising linguistic accommodation. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics*, pages 455–465.

Ewa Kacewicz, James W Pennebaker, Matthew Davis, Moongee Jeon, and C Arthur. 2013. Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, 33(2):125–143.

SL Kline and JM Ceropski. 1984. Person-centered communication in medical practice. In J T Wood and G M Phillips, editors, *Human Decision-Making*, pages 120–141. SIU Press, Carbondale.

M Natale. 1975. Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, 32(5):790–804.

KG Niederhoffer and JW Pennebaker. 2002. Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, 21(4):337–360.

B Noble and R Fernández. 2015. Centre Stage: How Social Network Position Shapes Linguistic Coordination. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*.

Olutobi Owoputi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider, and Noah Smith. 2013. Improved Part-of-Speech Tagging for Online Conversational Text with Word Clusters. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 380–391.

JW Pennebaker, RJ Booth, and ME Francis. 2007. Linguistic Inquiry and Word Count: LIWC.

Martin J Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, 27(2):169–190.

David Reitter, Johanna D Moore, and Frank Keller. 2010. Priming of syntactic rules in task-oriented dialogue and spontaneous conversation.

R B van Baaren, R W Holland, B Steenaert, and Ad van Knippenberg. 2003. Mimicry for money: Behavioral consequences of imitation. *Journal of Experimental Social Psychology*, 39(4):393–398.

Yafei Wang, David Reitter, and John Yen. 2014. Linguistic Adaptation in Conversation Threads: Analyzing Alignment in Online Health Communities. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*.

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.

A Supplemental Material

(Temporary collection of material that I'm considering adding as supplemental material – separate from the paper, doesn't count against 8-page limit. [references also don't count against limit])

[Proof that LLA and Healey's measure of lexical similarity can depend on message/reply length.]

Message	{∅}	{X}	{Y}	{X,Y}
{∅}	160	20	10	10
{X}	140	30	20	10
{Y}	170	10	15	5

Table 4: Another case where (mean lexical) ¿ categorical alignment.