Influence of Gender on Linguistic Alignment in Dialogues

Eui Yeon Jang
12661635

Hannah Lim
10588973

Daniël Nobbe
12891517

Abstract

Linguistic alignment is the tendency of changing word choices by a speaker to align to their partner. Previous work has found that this linguistic alignment is influenced by social factors however, little research has been done on the influence of gender on linguistic alignment. In this paper, we investigate the influence of gender dynamics on stylistic and lexical alignment in spoken multi-party dialogues in different domains. The results vary across corpora for stylistic alignment, where only males aligning more towards females in one corpus was found significant. Lexical alignment showed to be more consistent, where males have shown to align more towards females. Regarding the varying trends, we note the limitations of the available annotations of the corpora as well as various other features that play a role in measuring alignment.

1 Introduction

The Communication Accommodation Theory posits that interlocutors tend to coordinate their linguistic patterns with their dialogue partners in order to achieve certain social goals, such as enhancing social relations. Consequently, the social relations among interlocutors can be thought to influence the linguistic properties of a dialogue (Giles, 2008). Various research has demonstrated the influence of social power relations on linguistic alignment (Danescu-Niculescu-Mizil et al., 2012; Noble and Fernández, 2015; Doyle et al., 2016; Herring, 2008).

Research in sociolinguistics and dialogue modelling focused on the effect of gender in linguistic behaviour suggest a power dynamic between genders (Eecke and Fernández, 2016; Hemphill and Otterbacher, 2012). There is research into how gender dynamics affects linguistic alignment. For example, Bilous and Krauss (1988) and Mulac et al. (2006) have investigated whether there is a difference in non-lexical alignment between males and females towards the same gender or the opposite gender. Sabater (2017) studies alignment in online support groups which show that female participants align more towards male participants. However, there is little on how gender dynamics can affect linguistic alignment and whether differences in gender influence linguistic alignment in spoken dialogue.

In this paper, we explore how the gender of interlocutors influences the alignment within spoken multi-party dialogues. Specifically, we examine the lexical and stylistic alignment between genders using generalised linear models. In addition to various research on the power dynamic between genders and the influence of social power on alignment, we hypothesise that gender of interlocutors influences the level of linguistic alignment.

The study of relationship between linguistic coordination and gender can be valuable, as it can help shine light onto the gender dynamics, the social statuses associated with gender, and its influence in dialogue. Furthering knowledge on these subjects can help ensure fair and gender-neutral debating, negotiation and decision-making processes in society at large.

In the subsequent sections, we introduce related work on linguistic alignment and roles of gender in dialogues. In Section 3 and 4, we describe the data and methodology of our analysis. Section 5 and 6 presents and discusses the results.

2 Related Work

2.1 Gender and Alignment

Linguistic alignment in human conversation is the tendency to adapt to the language of conversational partners. There are multiple theories that explain the cause of this adaption. In this research we mainly focus on the Communication Accommodation Theory proposed by Giles (2008), which posits that speakers' alignment with their interlocutors is driven by social goals. The level of alignment is dependent on multiple factors, including the social power difference between the interlocutors. Danescu-Niculescu-Mizil et al. (2012) and Noble and Fernández (2015) have found that people tend to align more to interlocutors with a high or higher social status.

Another factor potentially affecting alignment between interlocutors is their gender. Previous works have investigated the effect of gender on various linguistic behaviours. However, at the time of writing, little research (Sabater, 2017) has gone into the effect of gender on the level of alignment in (spoken or written) dialogue.

Conversational behaviour of males and females tend to differ significantly on multiple levels. For instance, males tend to use different words than females (Koppel et al., 2002) which could even lead to predicting the gender of a speaker based solely on the words used in the utterance (Schofield and Mehr, 2016). Moreover, female speakers tend to be interrupted more often, and male speakers tend to interrupt female speakers more often than other male speakers. Male speakers mainly use interruptions to grab the floor, which is a sign of dominance (Eecke and Fernández, 2016), so it might be inferred that male speakers tend to behave more dominant. This could affect the social status of the male speaker, therefore affecting the level of alignment.

2.2 Alignment Models

There are several methods that look for presence of linguistic markers in utterances to compute alignment. Danescu-Niculescu-Mizil et al. (2012) implemented a subtractive conditional probability model, where the alignment is quantified as the difference between likelihood of a marker-containing utterance and a marker-containing reply. This measure, while simple and intuitive to compute, is influenced by the overall frequency of markers. The Hierarchical Alignment Model (HAM) proposed by Doyle et al. (2016) is an improvement of the subtractive model as it satisfies the following aspects: separability across markers, consistency across varying marker frequencies, and robustness to sparse data. HAM defines alignment as an additive effect in log-odds space. Neither of these models are robust to the length of an utterance, as they solely take into account the presence of a marker, not its count.

Level of Education	Female	Male
Bachelor of Arts	2	0
Other Bachelors	2	2
Masters	3	16
PhD	6	21
Postdoc	1	0
Professor	2	5

Table 1: Level of education of female and male speakers in the ICSI Meeting Corpus.

The Word-Based Hierarchical Alignment Model (WHAM) proposed by Doyle and Frank (2016) extends HAM by using marker token frequencies, thereby gaining robustness to utterance length.

Xu et al. (2018) proposed a simple generalised linear model which show that low-level linguistic features such as utterance lengths have significant effect on alignment, and that the effect of social power is not detected by the model when the length is corrected for. They find that there is an effect of power on sequence length, which in turn affects alignment. We adopt this approach to analyse linguistic alignment.

3 Data

3.1 Corpora

For our experiment, we use two datasets of spoken multi-party dialogue: ICSI Meeting Corpus (Janin et al., 2003), and AMI Meeting Corpus (Carletta et al., 2005).

The ICSI Meeting Corpus contains meetings recorded at the International Computer Science Institute at UC Berkeley, and as such contain dialogues with a general computer science theme. The corpus contains 75 recorded meetings with 61 unique speakers, on average 6 per meeting. The meetings have been transcribed and annotated with information about the speakers, dialogue acts, and the adjacency pairs of utterances and their replies. The corpus has 61 unique speakers - 16 female, 44 male, and one unspecified. It is also annotated with the level of education of each speaker, such as undergraduate or professor. In Table 1, the number of speakers corresponding to each education level is listed for each gender.

The AMI Meeting Corpus consists of recorded meetings from three European research institutes. One-third of the corpus is natural, uncontrolled meetings, and the rest is meetings with staged

Role	Female	Male
Marketing Expert	52	86
User Interface Designer	44	94
Industrial Designer	36	102
Project Manager	44	94

Table 2: Roles assigned to female and male participants in the AMI Meeting Corpus.

premises and participants playing different roles in a product design team. The natural meetings generally have a linguistics theme, and the staged meetings focus on the design of a television remote control. The corpus consists of a total of 138 recorded meetings, with 60 female and 130 male speakers. The speakers are generally non-native English speakers. Similar to the ICSI Meeting Corpus, a large amount of annotation information is available for the transcriptions, including adjacency pairs. In the staged meetings, annotation is available for the roles of speakers listed in Table 2. Note that the same speakers can play different roles. For the natural meetings, no such annotation is available. Therefore, for this research we only used the staged meetings.

3.2 Adjacency Pairs

To facilitate analysis, we only extract utterances annotated with information about their adjacency pairs. We refer to the initial utterance in the pair as the *prime* and its response, the *target*. In cases where multiple replies are given tot he same initial utterance, we create an adjacency pair for each reply using the initial utterance as *prime*.

For each utterance in an adjacency pair, the gender of the speaker of the utterance and a count of all words in the utterance is kept. Before counting, the words are pre-processed: stemmed by NLTK's dictionary-based stemmer, WordNet Lemmatizer (Bird et al., 2009); punctuation marks removed, and all letters lower-cased.

The set of processed adjacency pairs can be distinguished with respect to the gender of the *prime* speaker as well as *target* speakers, and thus consists of four basic subsets of adjacency pairs – male-to-male, male-to-female, female-to-male, and female-to-female. For more information on the implementation, please refer to our GitHub repository at `https://github.com/EuiYeonJang/CDM`.

Category	Examples
articles	a, an, the
pronouns	I, himself, somebody
prepositions	with, above, off
negations	no, never, neither
tentatives	maybe, anyone, guess
certainties	always, never
discrepancies	should, would, could
exclusions	but, except, without
inclusions	with, and, include

Table 3: Examples of words in LIWC categories.

4 Methodology

4.1 Linguistic Markers

We define alignment from *target* to *prime* to mean the use of a linguistic marker in *target* given its use in the *prime*. We consider different types of linguistic markers to measure stylistic and lexical alignment. As with Danescu-Niculescu-Mizil et al. (2012) and Xu et al. (2018), we consider the linguistic style of a speaker by their use of function words. Thus, interlocutors are said to align stylistically when the speaker uses words in the same function class in their reply as the *prime* speaker. The classes of function words are formalised by the LIWC (Pennebaker et al., 2015) word categories. We employ nine of the LIWC categories listed in Table 3, containing a total of 326 lexemes. The words from each category are considered *markers* for the category.

Lexical alignment looks at the specific words – an interlocutor aligns lexically when they use the same specific words in their reply as the original speaker. We take a similar approach with markers to measure lexical alignment, where each word is considered to be its own category, and therefore, are considered a marker for itself. We refer to the markers for the LIWC categories as stylistic markers, and the unique words as lexical markers.

4.2 Analytical Models

To measure linguistic alignment, we adopt the generalised linear model (GLM) proposed by Xu et al. (2018). It models the likelihood of a linguistic marker appearing in a *target* given the *prime* as having a Bernoulli distribution using the logit link function. Fitting a GLM with predictor terms representing linguistic features, the β values should indicate the influence of the features in predicting alignment.

prime (female)	But do *you* think *we* can simply take logit to be 1 if a **marker** occurs and 0 if *it* doesn't?			
target (male)	I think it's a start.			
	C_{count}	C_{gender}	C_{pLen}	logit
pronoun	3	0	21	1
'marker'	1	0	21	0

Table 4: Example adjacency pair, with C values and logit values specified for a linguistic marker (pronoun, *italic*) and a lexical marker ('marker', **bold**)

While we are interested in the influence of gender on alignment, as pointed by Xu et al. (2018), the effect observed may not be reliable when considered by itself. Therefore, we append the length of *prime* utterance as an additional predictor term. Equation 1 formalises this model. C_{count} represents the number of occurrences of a given linguistic marker in the *prime* utterance. C_{gender} is a binary predictor denoting the gender of the *prime* speaker, and C_{pLen} is the length of the *prime* utterance in number of words. Then, the terms such as $C_{\text{count}} * C_{\text{gender}}$ and $C_{\text{count}} * C_{\text{pLen}}$ characterises the influence of gender and length on alignment, respectively. Table 4 gives an example of how the C values and logit are determined.

$$
\begin{aligned}
\text{logit}(m) = \beta_0 &+ \beta_1 C_{\text{count}} + \beta_2 C_{\text{gender}} + \beta_3 C_{\text{pLen}} \\
&+ \beta_4 C_{\text{count}} * C_{\text{gender}} \\
&+ \beta_5 C_{\text{count}} * C_{\text{pLen}} \\
&+ \beta_6 C_{\text{gender}} * C_{\text{pLen}} \\
&+ \beta_7 C_{\text{count}} * C_{\text{gender}} * C_{\text{pLen}}
\end{aligned} \tag{1}
$$

We can also look at the strength of alignment towards the *prime* speaker by fitting a model with only C_{count} as predictor (Equation 2). Comparing the β_1 coefficient among the gender of *prime* speakers could give more insights on the influence of the *prime's* gender on alignment.

$$
\text{logit}(m) = \beta_0 + \beta_1 C_{\text{count}} \tag{2}
$$

4.3 Experiment Setup

We perform separate experiments for each equation introduced above[1] Each adjacency pair with respect to *each linguistic marker* is considered a data point. The *prime* utterance of the pair delivers the values for C_{count}, C_{gender}, C_{pLen} and all their combinations for a given marker. The *target* utterance then delivers a binary value for whether the marker is present in the utterance. A example is illustrated in Table 4. We fit the model to these data points, and retrieve Wald's z-scores and their p-values for each of coefficients of the fit model. This indicates how significant the predictor terms are to the model.

We perform the following experiments on each corpus and for each type of alignment.

Experiment 1. In this experiment, we are interested in how gender of the *prime* speaker and its length influences the alignment of *target* speaker. We group the adjacency pairs into two sets based on the gender *target* speaker, and fit the model from Equation 1 separately to each set of adjacency pairs. Examining the model's fit to $C_{\text{count}} * C_{\text{gender}}$ should indicate the influence of gender of *prime* speaker on alignment. Moreover, examining $C_{\text{count}} * C_{\text{pLen}}$ and $C_{\text{gender}} * C_{\text{pLen}}$ should give an interesting insight to how gender of the speaker may indirectly affect alignment through utterance length.

For further insights how the coefficient $C_{\text{count}} * C_{\text{gender}}$ is dependent on the utterance length of the *prime*, C_{pLen}, we adopt the same approach as Xu et al. (2018) by removing the interaction terms, $C_{\text{count}} * C_{\text{gender}} * C_{\text{pLen}}$, $C_{\text{count}} * C_{\text{pLen}}$ and $C_{\text{gender}} * C_{\text{pLen}}$, in a stepwise manner. Then, we compute the Akaike information criterion (AIC) score (Akaike, 1998) to examine and compare the quality of the models from which the interaction terms are removed. These results could support the findings of Xu et al. (2018), that neglecting low-level features such as utterance length of the *prime* could result in unreliable significant factors.

Experiment 2. With Equation 2, we investigate how each gender aligns to partners of a specific gender. In this experiment, we fit the equation for each of the four subsets of the adjacency pairs as introduced in Section 3. This shows whether the absolute level of alignment depends on the combination of the genders of the *prime* and *target* speakers. Therefore, if Experiment 1 has found that gender influences alignment, comparing the β_1 coefficients between the genders of *prime* speakers of this experiment will give further insights into the different strengths of alignment towards the *prime*

[1]We make use of the `python statsmodels v0.11.0 GLM` library using the default parameters. For implementation details, please refer to our GitHub repository.

Predictor	z-scores			
	ICSI		AMI	
	male	female	male	female
intercept	**-83.01******	**-71.18******	**-56.56******	**-76.42******
C_{count}	**31.55******	**25.59******	**16.10******	**22.61******
C_{gender}	**3.18****	1.35	**3.25****	**-6.18******
C_{pLen}	**-6.47******	**-7.63******	**-3.45****	**6.29******
$C_{\text{count}} * C_{\text{gender}}$	-0.82	-1.93	**-2.14***	-0.99
$C_{\text{count}} * C_{\text{pLen}}$	**-8.82******	**-7.82******	**-7.61******	**-10.19******
$C_{\text{gender}} * C_{\text{pLen}}$	-1.35	**2.93****	0.04	-0.80
$C_{\text{count}} * C_{\text{gender}} * C_{\text{pLen}}$	0.09	1.42	1.68	-0.43

Table 5: Results for stylistic alignment through Equation 1 (Experiment 1) on adjacency pairs split on gender of *target* speaker. Wald's z-score and significance level (*** for $p < 0.001$, ** for $p < 0.01$, * for $p < 0.05$).

genders. This implies that if it has been found that gender is not a significant predictor for alignment, the differences in β_1 should not be very salient.

5 Stylistic Alignment Analysis

Experiment 1. The results of Experiment 1 for stylistic alignment over all linguistic markers are summarised in Table 5. The effects of C_{count} are significant for both corpora, showing alignment of *target* to *prime*.

The effects of term $C_{\text{count}} * C_{\text{gender}}$ differs for each corpus. In the ICSI Meeting Corpus, the term is not significant for either *target* gender, and it is only significant for *target* gender male. This indicates that gender of the *prime* speaker doesn't influence alignment in the ICSI Corpus, while it does influence how males align in the AMI Corpus. Additionally, we observe that $C_{\text{count}} * C_{\text{pLen}}$ is significant over all *target* genders for both corpora. This shows that *prime* utterance length is a greater influence on alignment overall.

Table 6 shows the results for $C_{\text{count}} * C_{\text{gender}}$ for when the model is fit for each category of markers. In these results, we observe that the term is significant for some categories for male and female in the ICSI Corpus. This is in line with findings from Xu et al. (2018), where alignment in certain categories of words are sensitive to gender, however, the overall effect is neutralised.

We can draw additional insights from examining the β values of Tables 5 and 6.[2] When the values are negative, this indicates a decrease in the logit value when *prime* speaker is male ($C_{\text{gender}} = 1$). For *prime* speaker female ($C_{gender} = 0$), so a negative β_3 value increases the predicted probability of alignment, relative to the male *prime* speaker

[2] A negative z-score corresponds to a negative β value.

marker	z-scores			
	ICSI		AMI	
	male	female	male	female
articles	-1.18	1.60	**-2.047***	0.436
pronoun	-0.24	**-2.16***	0.197	0.504
prepositions	-0.37	-0.87	-0.574	0.275
negations	-0.77	**-2.54**	-1.132	1.123
tentative	-0.93	-1.28	0.838	-0.843
certainty	0.41	0.38	1.809	-0.455
discrepancy	-0.18	-0.38	0.386	0.350
exclusive	-0.53	0.65	-0.536	-0.249
inclusive	**-2.25***	**-2.164***	**-3.024****	-0.424

Table 6: Results of term $C_{\text{count}} * C_{\text{gender}}$ for stylistic alignment on Equation 1 per LIWC category, split on gender of target speaker. Wald's z-score and significance level (*** for $p < 0.001$, ** for $p < 0.01$, * for $p < 0.05$).

case. With this, we see that overall, males in the AMI Corpus tend to align more towards female than male *prime* speakers. This can also observed in Experiment 2.

Moreover, we look at how $C_{\text{count}} * C_{\text{gender}}$ is dependent on *prime* utterance length by stepwisely removing interaction terms containing C_{pLen}. Table 7 shows the z-scores of $C_{\text{count}} * C_{\text{gender}}$ and the AIC score of the remainder models - model after removing interaction terms. We observe that for the AMI Corpus, $C_{\text{count}} * C_{\text{gender}}$ is dependent on C_{pLen} for female *target* speakers but not male.

For male *target* of the ICSI Meeting corpus removing the interaction terms $C_{\text{count}} * C_{\text{gender}} * C_{\text{pLen}}$ and $C_{\text{gender}} * C_{\text{pLen}}$ results in a significant $C_{\text{count}} * C_{\text{gender}}$ coefficient. For the female *target* of the ICSI Meeting corpus the model that seems to have the highest quality is the full model or the model where the interaction term $C_{\text{count}} * C_{\text{gender}} * C_{\text{pLen}}$ is removed. It is notable that for every group that contains one model with a significant $C_{\text{count}} * C_{\text{gender}}$ coefficient, the model with the highest quality also has a significant z-score for the $C_{\text{count}} * C_{\text{gender}}$ coefficient.

Experiment 2. The β_1 values for stylistic alignment as part of Experiment 2 are reported in Table 8. The effects of C_{count} are significant for all groups in both corpora, capturing that there is stylistic alignment between all gender pairings.

The differences between the values between gender pairs in the ICSI Corpus are minute, which is in line with our observations during Experiment 1, that gender does not influence alignment.

For the AMI corpus, there are somewhat larger

Predictor	ICSI				AMI			
	male		female		male		female	
	z-scores	AIC	z-scores	AIC	z-scores	AIC	z-scores	AIC
Full	-0.819	326050	-1.929	**119759**	**-2.138***	**120021**	-0.986	46022
Full $-C_{\text{count}} * C_{\text{gender}} * C_{\text{pLen}}$	-1.079	**326048**	-1.311	**119759**	-1.3268	120022	**-2.650****	**46020**
Full $-C_{\text{count}} * C_{\text{gender}} * C_{\text{pLen}} - C_{\text{count}} * C_{\text{pLen}}$	-0.141	326296	-0.949	119875	-1.941	120396	**-3.051****	46197
Full $-C_{\text{count}} * C_{\text{gender}} * C_{\text{pLen}} - C_{\text{gender}} * C_{\text{pLen}}$	**-2.336***	**326048**	1.117	119773	-0.979	**120021**	**-3.508*****	**46020**
Full $-C_{\text{count}} * C_{\text{gender}} * C_{\text{pLen}} - C_{\text{count}} * C_{\text{pLen}} - C_{\text{gender}} * C_{\text{pLen}}$	-0.523	326295	1.871	119892	-1.467	120396	**-4.237*****	46197

Table 7: Wald's z-score and significance level (*** for $p < 0.001$, ** for $p < 0.01$, * for $p < 0.05$) of the $C_{\text{count}} * C_{\text{gender}}$ term, and the AIC scores of the remainder model after removing other interaction terms from the full model stepwisely for stylistic alignment on Equation 1 split on gender of *target* speaker.

Gender Pairs	ICSI		AMI	
	β_1	SE	β_1	SE
male-to-male	0.22	0.003	0.38	0.021
male-to-female	0.21	0.006	0.41	0.008
female-to-male	0.23	0.006	0.33	0.048
female-to-female	0.21	0.007	0.43	0.019

Table 8: β_1 values and standard error for C_{count} for stylistic alignment through Equation 2 (Experiment 2) for each *target*-to-*prime* gender pairs. All values are significant with $p < 0.01$ according to Wald's z-score.

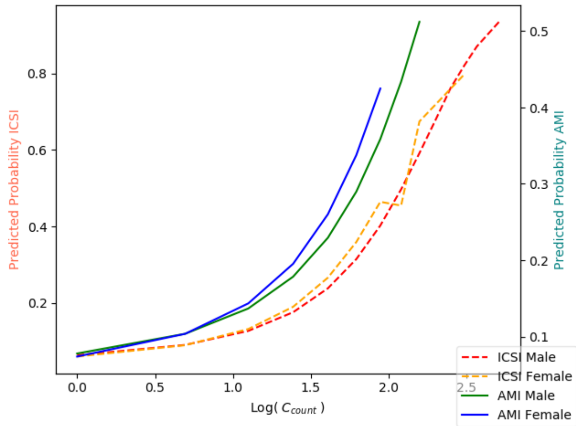

Figure 1: On the vertical axis, the median predicted probability (inverse logit) as given by our lexical alignment model for experiment 2, per prime gender and dataset. On the horizontal axis, the log of C_{count}. This indicates how likely lexical alignment is to happen. There is a clear difference between the genders.

differences, and we see the general pattern as Experiment 1: there is more alignment towards female *prime* speakers. Male *target* speakers show less varying behaviour, whereas for female *target* speakers there is a larger difference between the *prime* genders.

6 Lexical Alignment Analysis

Experiment 1. The results of Experiment 1 on lexical alignment are summarised in Table 9. As indicated by C_{count}, there is alignment by all *target* speakers in both corpora. We see the effect of $C_{\text{count}} * C_{\text{gender}}$ is only significant for male *target* speakers for both corpora.

In the ICSI corpus, the β_1 value for male *target* negative. This indicates that males align more towards females than towards other males, a similar result to the stylistic alignment analysis. For the AMI corpus, however, the β_1 value for *prime* male is positive, implying that males tend to align more towards other males than females in this corpus. This is the opposite of the behaviour seen in stylistic alignment. This indicates there is not only a difference in general alignment behaviour between the two gender groups, but also a difference in the type of alignment.

We again see in Table 11 that $C_{\text{count}} * C_{\text{gender}}$ become significant when removing C_{pLen} terms for female *target* speakers for both corpora. Whereas, the term is significant significant in the full model, which is also the model with highest quality according to AIC scores. This could indicate that $C_{\text{count}} * C_{\text{gender}}$ for male *target* speakers is not or less dependent on the utterance length of the *prime* speaker.

Experiment 2. Table 10 reports the results for Experiment 2 on lexical alignment. All these β_1 values are negative, which indicates that lexical alignment is negative - a term that occurs more

Predictor	z-scores			
	ICSI		AMI	
	male	female	male	female
intercept	**131.31*****	**104.65*****	**43.51*****	**64.15*****
C_{count}	**-182.43*****	**-147.39*****	**-53.79*****	**-83.57*****
C_{gender}	**15.68*****	-0.76	**-5.27*****	-0.30
C_{pLen}	**-67.92*****	**-58.77*****	**-22.58*****	**-34.84*****
$C_{\text{count}} * C_{\text{gender}}$	**-17.77*****	1.29	**4.87*****	-1.64
$C_{\text{count}} * C_{\text{pLen}}$	**79.81*****	**66.36*****	**17.18*****	**33.60*****
$C_{\text{gender}} * C_{\text{pLen}}$	**-16.14*****	-0.59	**2.06***	-1.37
$C_{\text{count}} * C_{\text{gender}} * C_{\text{pLen}}$	**25.23*****	**5.11*****	0.768	0.13

Table 9: Results for lexical alignment through Equation 1, split on gender of *target* speaker. Wald's z-score and significance level (*** for $p < 0.001$, ** for $p < 0.01$, * for $p < 0.05$).

Gender Pairs	ICSI		AMI	
	β_1	SE	β_1	SE
male-to-male	-3.66	0.017	-3.83	0.055
male-to-female	-3.41	0.009	-4.52	0.018
female-to-male	-3.56	0.022	-4.37	0.036
female-to-female	-3.80	0.017	-4.21	0.049

Table 10: β_1 values and stadard error for C_{count} for lexical alignment through Equation 2 (Experiment 2) for each *target*-to-*prime* gender pairs. All values are significant with $p < 0.01$ according to Wald's z-score.

in *prime* is less likely to appear in *target*. This is in line with the results from experiment 1, in table 9, the z-scores for the C_{count} term are also negative. Since the lexical alignment model examines *all* words in the dataset, it is likely that many of the *prime* words do not appear in *target* at all, regardless of their count. As a result, the β_1 values themselves have little meaning. However, the differences between them do have a meaning.

In the ICSI corpus, *target* speakers align more to the opposite gender, as the β_1 is higher. In the AMI dataset, on the other hand, more lexical alignment happens when both *target* and *prime* are the same gender. The difference is especially salient for male *target* speakers.

Figure 1 shows a plot of the predicted probability for this analysis, over the log of C_{count} values. The utterance pairs are split between prime gender for each dataset. A distance between these lines indicates that there is a different alignment behaviour towards the two genders. We see that in both datasets, the female alignment curve is slightly higher, indicating that speakers are more likely to align to female partners.

7 Discussion and Conclusion

For overall stylistic alignment, we only find a significant effect of gender on alignment for male *target* speakers in the AMI Corpus, where they align more towards females than males. Female *target* speakers showed no significant effect.

We find from Experiment 2 that more stylistic alignment is present towards the female *prime* speakers for both *target* genders in the AMI Corpus. In the ICSI Corpus, there were no significant differences in stylistic alignment between *prime-target* pairs for Experiment 2.

However, it is interesting to note that, although we do not observe that gender influences alignment for female *target* speakers, we see that the $C_{\text{count}} * C_{\text{pLen}}$ and $C_{\text{gender}} * C_{\text{pLen}}$ are significant. This could suggest that the gender has an indirect influence on alignment via the length of utterance. This could possibly explain the comparatively greater difference between female-to-male and female-to-female values as opposed to male-to-male and male-to-female values in Table 8.

In lexical alignment, we find that the effect of gender on alignment is significant for male *target* speakers for both corpora. In the ICSI Corpus, male *target*s align more towards female *prime*s, whereas in the AMI Corpus, male *target*s align more towards male *primes*. Female *target* speakers show no significance for either corpora.

From Experiment 2, we find that there is more alignment towards the opposite gender in the ICSI Corpus, whereas in the AMI Corpus, there is more alignment towards the same gender.

From Figure 1, alignment towards female *prime* speakers is somewhat higher.

From these results, we see in general that there is more often a significant effect of the *prime* gender on their alignment for male *target* speakers. As we observe significance of other predictor terms for female *target* speakers, we conclude that the lack of influence of gender for female *target*s is not due to the availability of data for females. This means that the alignment of males is *in general* affected by the gender of their dialogue partners. There are some differences between how this effect is expressed between the corpora and the types of alignment we examined. Females also show alignment behaviour that depends on the gender of their interlocutor, but this is less significant. As the findings of Table 11 seem to suggest that removing the interactions terms containing the utterance length of the *prime* speaker results in a significant effect for the female *target* speakers, we can conclude that neglecting the utterance length would have led to unreliable results.

There are notable differences in the males' lexical alignment behaviour with respect to gender between the two corpora, whereas stylistic alignment shows similar trends. In the ICSI Corpus, males tend to lexically align more towards females, whereas in the AMI, males align more towards males. As the statistic on the power status of the ICSI Corpus Table 1, males are more often in a position of power, there appears to be another process at work. For the lexical alignment in the

Predictor	ICSI				AMI			
	male		female		male		female	
	z-scores	AIC	z-scores	AIC	z-scores	AIC	z-scores	AIC
Full	**-17.767*****	**515063**	1.291	**184258**	**4.873*****	161213	-1.637	60621
Full $-C_{\text{count}} * C_{\text{gender}} * C_{\text{pLen}}$	0.931	515658	**8.330*****	184283	**10.347*****	**161211**	**-2.909****	**60619**
Full $-C_{\text{count}} * C_{\text{gender}} * C_{\text{pLen}} - C_{\text{count}} * C_{\text{pLen}}$	**-13.603*****	542351	**7.336*****	195088	**11.767*****	164297	**-2.595****	61894
Full $-C_{\text{count}} * C_{\text{gender}} * C_{\text{pLen}} - C_{\text{gender}} * C_{\text{pLen}}$	1.795	515700	**9.084*****	184323	**9.881*****	161251	**-2.724****	60624
Full $-C_{\text{count}} * C_{\text{gender}} * C_{\text{pLen}} - C_{\text{count}} * C_{\text{pLen}} - C_{\text{gender}} * C_{\text{pLen}}$	**-13.617*****	542350	**8.076*****	195104	**11.804*****	164315	**-2.707****	61900

Table 11: Wald's z-score and significance level (*** for $p < 0.001$, ** for $p < 0.01$, * for $p < 0.05$) of the $C_{\text{count}} * C_{\text{gender}}$ term, and the AIC scores of the remainder model after removing other interaction terms from the full model stepwisely for lexical alignment on Equation 1 split on gender of *target* speaker.

AMI Corpus, we see that males tend to align more towards males. As generally males are in higher status functions than females, this could explain the lexical alignment for male *target* speaker. If this would be a contributing factor, however, one would expect female speakers to show the same patterns, which they do not.

Another potential difference between these two corpora is their location. ICSI Corpusis recorded fully in Berkely (USA), while the AMI Corpus is recorded in various locations in Europe. There are two potential effects of this. The ICSI Corpus has a majority of native English speakers, where the AMI Corpus is mostly made up of non-native English speakers, which could affect the alignment behaviour. Additionally, there could be cultural difference between the two corpora, affecting alignment behaviour.

We find some trends of the influence of gender on alignment in our analyses. However, we need to be careful to generalise our results. Despite having corpora from different domains, the majority of the dialogues in either corpora are quite specific (meetings at a computer science institute and product design meetings). Furthermore, there are various outside factors that can influence alignment, such as roles and levels of education of the interlocutors, their age, cultural backgrounds and the hierarchical relations. Unfortunately, the annotations for these features are not always available. It is a suggestion for future research to compensate for these various aspects to better understand the influence of gender dynamics on linguistic alignment.

Another possible limitation of our results to note is the difference in criteria for what is considered as adjacency pairs. In the ICSI Corpus, a dialogue act can be annotated as an adjacency pair even if it is a response to a question that opened a discussion. Adjacency pairs in the AMI Corpus, are generally closer together and are only annotated if they are direct responses. This means that some pairs are much further apart, which may dilute the effect of alignment that we attempt to observe. This may also explain the difference in results between the two corpora.

For future research, we would like to see some of the uncertainties in this work addressed. First of all, it would be interesting to see the results of other corpora. In the two corpora we examined, the settings were somewhat similar, but the results were not the same, especially with regards to stylistic alignment. Applying this analysis to more corpora could give a better idea of general human behaviour. This could also prove or disprove cultural effects on alignment behaviour with respect to gender.
Additionally, as mentioned above, it would be interesting to include power status in this analysis, to decouple those effects from gender. This could be done by including a C_{power} term and its interaction terms in equation 1. The ICSI Corpus could be a viable dataset for this analysis, but another dataset with binary power levels, such as the one used by Xu et al. (2018), could be even more suitable.

Conclusion In this work, we apply generalised linear models to two multi-party dialogue corpora, to find how stylistic and lexical alignment depend on the gender of both interlocutors in an adjacency pair.

For overall stylistic alignment, we find that in one of the corpora, the AMI Corpus, male speakers align more towards female speakers. In the other corpus, there is no significant effect of gender on overall stylistic alignment. For lexical alignment, we find that the alignment behaviour of male speakers is dependent on the gender of their interlocutor, whereas the alignment behaviour of female speakers is not. Curiously, male speakers align more towards other male speakers in the AMI Corpus, and more towards female speakers in the ICSI Corpus.

We find that the corpora show marked differences, and propose some reasons as to why this is, and how our analysis could be expanded on.

Acknowledgments

The authors equally contributed to this study. The listing order is random. Daniel focused on preprocessing the two datasets. Eui Yeon did the lexical and stylistic analysis. Hannah analysed the dependence on the utterance length. Analysis and discussion of the results was done with equal contribution from all authors.

References

Hirotogu Akaike. 1998. Information theory and an extension of the maximum likelihood principle. In *Parzen E., Tanabe K., Kitagawa G. (eds) Selected Papers of Hirotugu Akaike. Springer Series in Statistics (Perspectives in Statistics).*

Frances R. Bilous and Robert M. Krauss. 1988. Dominance and accomodation in the conversational behaviours of same- and mixed-gender dyads. *Language and Communication*, 8(3):183–194.

Steven Bird, Edward Loper, and Ewan Klein. 2009. *Natural Language Processing with Python*. O'Reilly, Sebastopol, California.

Jean Carletta, Simone Ashby, Sebastien Bourban, Mike Flynn, Thomas Franz Hain Mael Guillemot, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa Kronenthal, Iain McCowan, Guillaume Lathoud, Mike Lincoln, Agnes Lisowska, Wilfried Post, Dennis Reidsma, and Pierre Wellner. 2005. The ami meeting corpus. In *Proceedings Measuring Behavior 2005, 5th International Conference on Methods and Techniques in Behavioral Research. L.P.J.J. Noldus, F. Grieco, L.W.S. Loijens and P.H. Zimmerman (Eds.), Wageningen: Noldus Information Technology*.

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 699–708, New York, NY, USA. Association for Computing Machinery.

Gabriel Doyle and Michael C Frank. 2016. Investigating the sources of linguistic alignment in conversation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 526–536.

Gabriel Doyle, Dan Yurovsky, and Michael C. Frank. 2016. A robust framework for estimating linguistic alignment in twitter conversations. In *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, page 637–648, Republic and Canton of Geneva, CHE. International World Wide Web Conferences Steering Committee.

Paul Van Eecke and Raquel Fernández. 2016. On the influence of gender on interruptions in multiparty dialogue. In *INTERSPEECH*.

H. Giles. 2008. Communication accommodation theory: "when in rome ..." or not!. In *Engaging theories in interpersonal communication: Multiple perspectives*, chapter 12, pages 161–174. L. A. Baxter and D. O. Braithwaite (Eds.).

Libby Hemphill and Jahna Otterbacher. 2012. Learning the lingo? gender, prestige and linguistic adaptation in review communities. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, page 305–314, New York, NY, USA. Association for Computing Machinery.

Susan C. Herring. 2008. Gender and power in on-line communication. In Janet Holmes and Miriam Meyerhoff, editors, *The Handbook of Language and Gender*, volume 25 of *Blackwell Handbooks in Linguistics*, chapter 9, pages 202–228. Wiley.

Adam Janin, Don Baron, Jane Edwards, Dan Ellis, David Gelbart, Nelson Morgan, Barbara Peskin, Thilo Pfau, Elizabeth Shriberg, Andreas Stolcke, and Chuck Wooters. 2003. The icsi meeting corpus. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03).*, volume 1, pages 364—-367.

Moshe Koppel, Shlomo Argamon, and Anat Rachel Shimoni. 2002. Automatically Categorizing Written Texts by Author Gender. *Literary and Linguistic Computing*, 17(4):401–412.

Anthony Mulac, Lisa B. Studley, John M. Wiemann, and James J. Bradac. 2006. Male/Female Gaze in Same-Sex and Mixed-Sex Dyads: Gender-Linked Differences and Mutual Influence. *Human Communication Research*, 13(3):323–343.

Bill Noble and Raquel Fernández. 2015. Centre stage: How social network position shapes linguistic coordination. In *Proceedings of the 6th Workshop on Cognitive Modeling and Computational Linguistics*, pages 29–38, Denver, Colorado. Association for Computational Linguistics.

James W. Pennebaker, Roger J. Booth, Ryan L. Boyd, and Martha E. Francis. 2015. Linguistic inquiry and word count: Liwc2015. *Austin, TX: Pennebaker Conglomerates (www. LIWC. net).*

Carmen P. Sabater. 2017. Linguistic accommodation in online communication: The role of language and gender. *Revista signos*, 50:265 – 286.

Alexandra Schofield and Leo Mehr. 2016. Gender-distinguishing features in film dialogue. In *Proceedings of the Fifth Workshop on Computational Linguistics for Literature*, pages 32–39, San Diego, California, USA. Association for Computational Linguistics.

Yang Xu, Jeremy Cole, and David Reitter. 2018. Not that much power: Linguistic alignment is influenced more by low-level linguistic features rather than social power. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 601–610, Melbourne, Australia. Association for Computational Linguistics.