Detecting Argumentative Discourse Acts with Linguistic Alignment

Timothy Niven and **Hung-Yu Kao**

Intelligent Knowledge Management Lab
Department of Computer Science and Information Engineering
National Cheng Kung University
Tainan, Taiwan
`tim.niven.public@gmail.com`, `hykao@mail.ncku.edu.tw`

Abstract

We report the results of preliminary investigations into the relationship between linguistic alignment and dialogical argumentation at the level of discourse acts. We annotated a proof of concept dataset with illocutions and transitions at the comment level based on Inference Anchoring Theory. We estimated linguistic alignment across discourse acts and found significant variation. Alignment features calculated at the dyad level are found to be useful for detecting a range of argumentative discourse acts.

1 Introduction

Argumentation mining remains a difficult problem for machines. Even for humans, understanding the substance of an argument can involve complex pragmatic interpretation (Cohen, 1987). Consider the reply of *B* in Figure 1. Absent broader conversational context, and perhaps knowledge of the background beliefs of *B*, it can be difficult to judge whether they are asking "which religions are correlated with increased life expectancy?" (pure questioning) or giving their opinion that "not just any religion is correlated with a longer life" (assertive questioning). Since only the latter is an argumentative discourse unit (ADU) (Stede, 2013), ambiguities like this therefore make it difficult to accurately identify the structure of argumentation.

In this work we investigate using a subtle yet robust signal to resolve such ambiguity: linguistic alignment. Alignment can be calculated in an unsupervised manner and does not require textual understanding. It is therefore well suited to our current technology as an extra pragmatic feature to assist dialogical argumentation mining. Our hypothesis is that, since alignment has been shown to relate to communication strategies (Doyle and Frank, 2016), different alignment effects will be

A: ...To be able to claim that life expectancy and health are tied to religion you have to rule out hundreds of other factors: diet; lifestyle; racial characteristics; genetic pre-disposition (religion tends to run in families) etc...

B: ...Can I just have ANY religion and have a longer life?

Figure 1: An example dyad from our dataset. Without disambiguating information it is hard to know if B's reply is pure or assertive questioning.
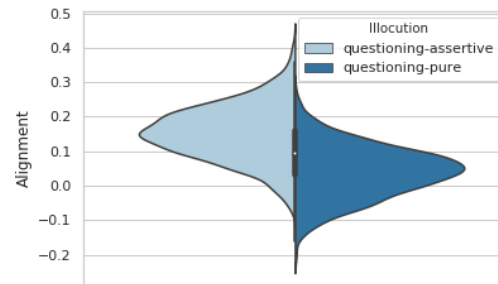

Figure 2: Posterior densities on alignment estimates for pure and assertive questioning in our dataset, indicating that alignment can help to disambiguate discourse acts.

observed over different argumentative discourse acts, providing signal for their detection. For example, Figure 2 shows our estimated posterior densities for alignment scores over pure and assertive questioning. On this basis, if *B*'s comment in Figure 1 is accompanied by a significantly positive alignment score, we would be correct more often than not classifying it as assertive questioning.

In this preliminary work we aim to address the following questions:

1. Are the majority of argumentative discourse acts associated with significantly different alignment effects?

2. Are alignment features useful for detecting argumentative discourse acts?

2 Background and Related Work

Linguistic alignment is a form of communication accommodation (Giles et al., 1991) whereby speakers adapt their word choice to match their interlocutor (Niederhoffer and Pennebaker, 2002). It can be calculated as an increase in the probability of using a word category having just heard it, relative to a baseline usage rate. An example is given in Figure 3. Note that alignment is calculated over *non-content* word categories.[1] While content words are clearly set by the topic of conversation, the usage rates of particular non-content word categories has shown to be a robust measure of linguistic style (Pennebaker and King, 2000). Consistent with previous work, we focus on alignment over the Linguistic Inquiry and Word Count (LIWC) categories (Pennebaker et al., 2015), listed in Table 1.

Linguistic alignment is a robust phenomenon found in a variety of settings. It has been used to predict employment outcomes (Srivastava et al., 2018), romantic matches (Ireland et al., 2011), and performance at cooperative tasks (Fusaroli et al., 2012; Kacewicz et al., 2014). People have been found to align to power (Willemyns et al., 1997; Gnisci, 2005; Danescu-Niculescu-Mizil et al., 2011), to people they like (Bilous and Krauss, 1988; Natale, 1975), to in-group members (Shin and Doyle, 2018), and to people more central in social networks (Noble and Fernandez, 2015). The variety of these contexts suggest alignment is ubiquitous and modulated by a complex range of factors.

Some previous work bears on argumentation. Binarized alignment features indicating the presence of words from LIWC categories were found to improve the detection of disagreement in online comments (Rosenthal and McKeown, 2015). We utilize more robust calculation methods that account for baseline usage rates which thereby avoid mistaking similarity for alignment (Doyle et al., 2016). Accommodation of *body movements* was found to decrease in face-to-face argumentative conflict where interlocutors had fundamentally differing opinions (Paxton and Dale, 2013; Duran and Fusaroli, 2017). In contrast we are concerned with linguistic forms of alignment.

A's message	B's reply	
	has pronoun	no pronoun
has pronoun	8	2
no pronoun	5	5

Figure 3: Example of linguistic alignment using a binarized "by-message" calculation technique (Doyle and Frank, 2016). B's baseline usage rate of pronouns is 0.5, coming from the bottom row. The top row shows the probability of B using a pronoun increases to 0.8 after seeing one in A's message.

Category	Examples	Usage
Article	*a, the*	0.076
Certainty	*always, never*	0.016
Conjunction	*but, and, though*	0.060
Discrepancy	*should, would*	0.018
Negation	*not, never*	0.018
Preposition	*to, in, by, from*	0.137
Pronoun	*it, you*	0.108
Quantifier	*few, many*	0.025
Tentative	*maybe, perhaps*	0.030
Insight	*think, know, consider*	0.027
Causation	*because, effect, hence*	0.021

Table 1: LIWC dictionary categories we use, examples, and baseline production rates observed in our dataset of ~ 1.5 million comments on news articles.

We focus on the argumentative discourse acts of Inference Anchoring Theory (IAT) (Budzynska and Reed, 2011; Budzynska et al., 2016). IAT is well motivated theoretically, providing a principled way to relate dialogue to argument structure. As noted above, an utterance that has the surface form of a question may have different functions in an argument - asking for a reason, stating a belief, or both. The IAT framework is designed to make these crucial distinctions, and covers a comprehensive range of argumentative discourse acts.

Two previous datasets are similar to ours. The US 2016 Election Reddit corpus (Visser et al., 2019) comes from our target genre and is reliably annotated with IAT conventions. However, the content is restricted to a single topic. Furthermore, political group effects have already been demonstrated to influence alignment (Shin and Doyle, 2018). These considerations limit our ability to generalize using this dataset alone. The Internet Argument Corpus (Abbott et al., 2016), used in prior work on disagreement (Rosenthal and McKeown, 2015), is much larger than our current dataset, however the annotations do not cover the principled and comprehensive set of discourse acts that we require to support dialogical argumentation mining in general.
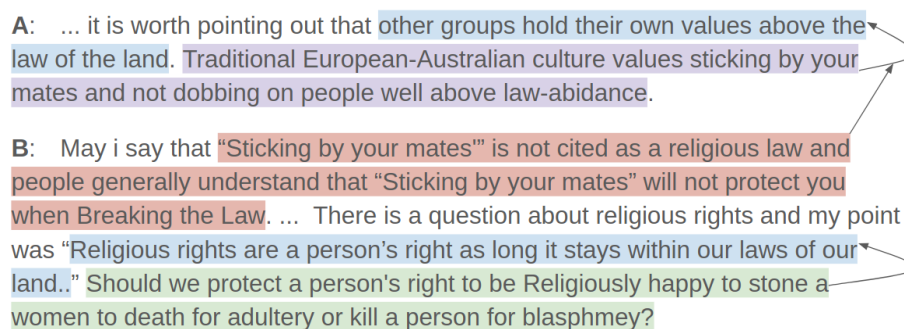
---

[1]Previous work has indicated the primacy of word-based over category-based alignment (Doyle and Frank, 2016). We leave investigation of alignment over words in argumentation to future work.

A: ... it is worth pointing out that other groups hold their own values above the law of the land. Traditional European-Australian culture values sticking by your mates and not dobbing on people well above law-abidance.

B: May i say that "Sticking by your mates'" is not cited as a religious law and people generally understand that "Sticking by your mates" will not protect you when Breaking the Law. ... There is a question about religious rights and my point was "Religious rights are a person's right as long it stays within our laws of our land.." Should we protect a person's right to be Religiously happy to stone a women to death for adultery or kill a person for blasphmey?

Figure 4: Annotating discourse acts across a message-reply pair. The blue text spans are *Asserting*. The red span is *Disagreeing*, which always crosses the comments - in this case attacking the inference in *A*. If *A* was the reply we would annotate the purple span as *Arguing*, as it offers a reason in support of the preceding assertion. In the reply, *Arguing* is provided by the green span, which is an instance of *Assertive Questioning*. Note that we only annotate what is in *B*. This pair is therefore annotated as: {*Asserting*, *Disagreeing*, *Assertive Questioning*, *Arguing*}.

3 Dataset

In this section we outline our annotation process. So far we have 800 message-reply pairs but annotated by just a single annotator. In future work we will scale up considerably with multiple annotators, and include Mandarin data for cross-linguistic comparison.

3.1 Source

We scraped ∼1.5M below the line comments from an academic news website, The Conversation,[2] covering all articles from its inception in 2011 to the end of 2017. In order to maximize the generalizabilty of our conclusions we selected comments covering a variety of topics. We also picked as evenly as possible from the continuum of controversiality, as measured by the proportion of deleted comments in each topic. More controversial topics are likely to see higher degrees of polarization, which should affect alignment across groups (Shin and Doyle, 2018). The most controversial topics we included are *climate change* and *immigration*. Among the least controversial are *agriculture* and *tax*.

Nevertheless this data source has its own peculiarities that attenuate liberal generalization. As the site is well moderated, comments are on topic and abusive comments are deleted, even if they also contain argumentative content. The messages are generally longer and less noisy than, for example, Twitter data. Moreover, many commenters are from research and academia. Therefore in general we see a high quality of writing, and of argumentation.

3.2 Annotation

The list of illocutions we chose to annotate are taken from Budzynska et al. (2016): *Asserting*, *Ironic Asserting*, *(Pure/Assertive/Rhetorical) Questioning*, *(Pure/Assertive/Rhetorical) Challenging*, *Conceding*, *Restating*, and *Non-Argumentative* (anything else). The transitions we consider follow IAT conventions. *Arguing* holds over two units, where a reason is offered as support for some proposition. *Disagreeing* occurs where an assertion conflicts with another. *Agreeing* is instantiated by phrases such as "I agree" and "Yeah."

Annotating *Rhetorical Questioning/Challenging* is the most difficult. As noted by Budzynska et al. (2016), there is no common specification for *Rhetorical Questioning*. We follow their definition, by which *Pure* and *Assertive Questioning/Challenging* ask for the speaker's opinion/evidence, and the *Assertive* and *Rhetorical* types communicate the speakers own opinion. Therefore the *Pure* varieties do not convey the speakers opinion, and the *Rhetorical* types do not expect a reply. Annotating *Rhetorical Questioning/Challenging* therefore requires a more complicated pragmatic judgment of the speaker's intention.

Our annotation scheme departs from previous work in that we only annotate at the comment and not the text segment level. Multiple annotations often apply to a single comment. An example is given in Figure 4. The text spans of the identified illocutions are highlighted and the transitions are indicated with arrows for clarity, but note that we did not annotate at that level.

Another difference from prior work relates to *Concessions*. Unlike Budzynska et al. (2016) we do not explicitly annotate the sub-type *Popular Concession* - where a speaker concedes in order to prepare the ground for disagreement. A potential confound with the annotation scheme described so far is ambiguous cases of *Agreeing* and *Disagreeing* in the same comment, which could be expected in a *Popular Concession*: "Yeah, I agree that X, but [counter-argument]." Because we are annotating at the level of the comment, we are able to distinguish these cases by considering combinations of discourse acts. A *Popular Concession* is distinguished by the presence of *Conceding* along with *Disagreeing*, optionally with *Agreeing*. A *Pure Concession* is then distinguished by the presence of *Conceding* and the absence of *Disagreeing*. We therefore do not need to rule that only one of *Agreeing* or *Disagreeing* can occur in a single comment.

We found that *Asserting* (627/800), *Arguing* (463/800), and *Disagreeing* (402/800) are by far the most common individually, and as a combination (339/800), reflecting the argumentative nature of our dataset. The distribution of comments over discourse acts is Zipfian. The lowest frequency discourse act is *Ironic Asserting*, which has only 12 annotations in our 800 comments.

4 Methodology

4.1 Alignment over Discourse Acts

To estimate alignment scores across discourse acts we parameterize the message and reply generation process as a hierarchy of normal distributions, following the word-based hierarchical alignment model (WHAM) (Doyle and Frank, 2016). Each message is treated as a bag of words and word category usage is modeled as a binomial draw. WHAM is based on the hierarchical alignment model (HAM) (Doyle et al., 2016), adapted by much other previous work (Doyle and Frank, 2016; Yurovsky et al., 2016; Doyle et al., 2017). WHAM's principal benefit over HAM is controlling for message length, which was shown to be important for accurate alignment calculation (Doyle and Frank, 2016). Our adaptation is shown in Figure 5. For further details of WHAM we refer the reader to the original work.

A key problem we need to address is our inability to aggregate counts over all messages in a conversation between two speakers (as in Figure 3).
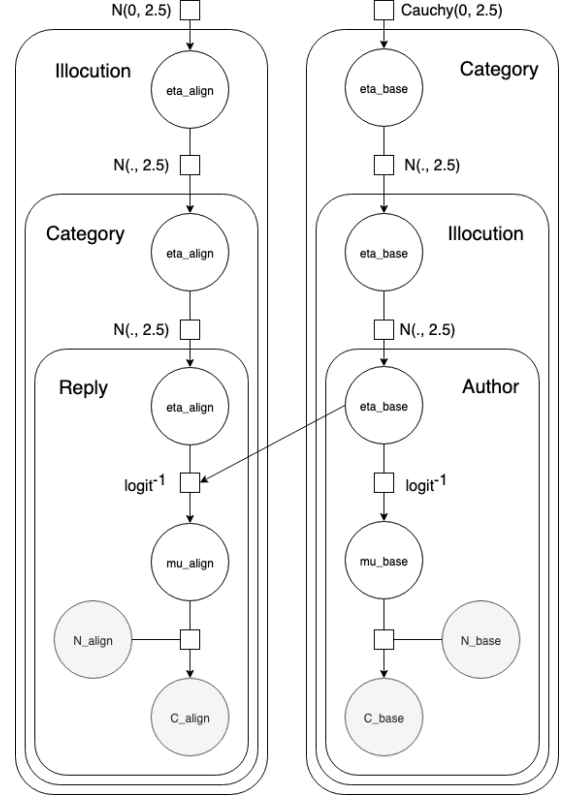

Figure 5: Our adaptation of WHAM (Doyle and Frank, 2016) for estimating alignment over argumentative discourse acts.

This is a virtue of the original WHAM model that provides more reliable alignment statistics. We cannot aggregate counts over multiple message-reply pairs since our target is the discourse acts in *individual* replies. However, we are helped somewhat by the long average comment length in our chosen genre ($\mu = 82.5$ words, $\sigma = 66.5$). The lowest baseline category usage rate is approximately 0.8% ($\mu = 3.6\%$, $\sigma = 2.2\%$). Therefore an average comment length gives us enough opportunity to see much of the effects of alignment on the binomial draw, but is likely to systematically underestimate alignment. In future work we will investigate this phenomenon with simulated data, and continue to search for a solution that makes better use of the statistics.

However, we can make more robust estimates of the baseline rate of word category usage by considering our entire dataset (~ 1.5 million comments). We have annotations for 261 authors. The most prolific author has $11,327$ comments. On average an author has 429 comments ($\sigma = 1,409$). For most authors we find multiple replies to comments that do not have each word category, making these statistics relatively reliable.
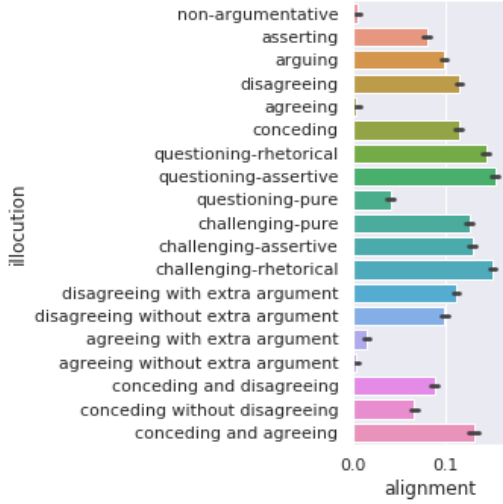
Figure 6: Alignment estimates over IAT discourse acts and combinations of interest. The error bars represent 95% highest posterior density.

Figure 7: ROC AUC Performance change from bag of GloVe vectors due to adding alignment features.

Bayesian posteriors for discourse act alignments are then estimated using Hamiltonian Monte Carlo, implemented with PyStan (Carpenter et al., 2017). We use $1,000$ iterations of No U-Turn Sampling, with 500 warmup iterations, and 3 chains. To address research question (1) we then compare the posterior densities of the last 500 samples from each chain, and look for significant differences in the means.

4.2 Alignment Over Comments

In this preliminary work, we use a simpler method for local alignment at the individual comment-reply level that we found effective. We utilize the author baselines calculated for each LIWC category from the entire dataset. Then, for each message and reply, we calculate the local change in logit space from the baseline to the observed usage rate, based on the binary criterion of whether the original message contained a word from the category. Formally, let the LIWC categories used in the first message be $\mathbb{C}_a$. For a LIWC category c, given the baseline logit space probability $\eta^{(c)}$ of the replier, and the observed usage rate r of words from category c in the reply, we calculate the alignment score as

$$ s^{(c)} = \begin{cases} \text{logit}(r) - \eta^{(c)} & c \in \mathbb{C}_a \\ 0 & \text{otherwise} \end{cases} $$

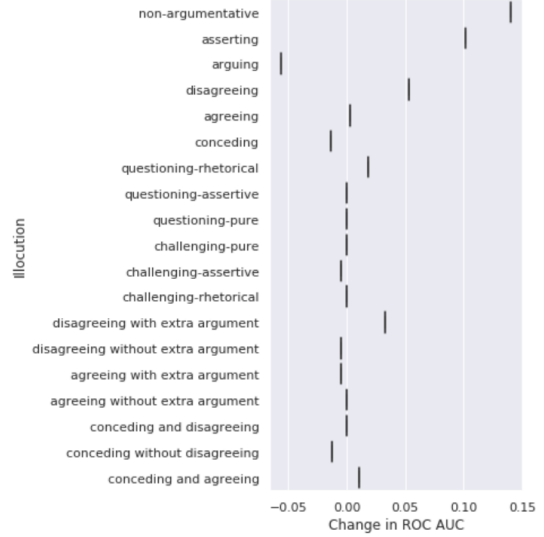We clip these values to be in the range $[-5, 5]$

to avoid infinite values and floor effects - for example where the reply does not contain a word from c. This range is large enough to cover the size of alignment effects we observed. Following this calculation method we end up with an 11-dimensional vector of alignments over each LIWC category for each reply.

4.3 Detecting Argumentative Discourse Acts

To investigate our second preliminary research question we perform logistic regression for each annotated comment and each discourse act. Our baseline is a bag of GloVe vectors (Pennington et al., 2014). We use the 25-dimensional vectors trained on 27 billion tokens from a Twitter corpus. We concatenate the 11-dimensional alignment score vector to the bag of GloVe representation and look for an increase in performance. We randomly split the dataset into 600 training data points, and 200 for testing. We implement logistic regression with Scikit-learn (Pedregosa et al., 2011) and use the LBFGS solver. We set the maximum number of iterations to $10,000$ to allow enough exploration time. Because this is not a deterministic algorithm, we take the mean performance of 20 runs over different random seeds as the final result. As we are concerned with detection, and because the labels in each class are very imbalanced, our evaluation metric is ROC AUC.
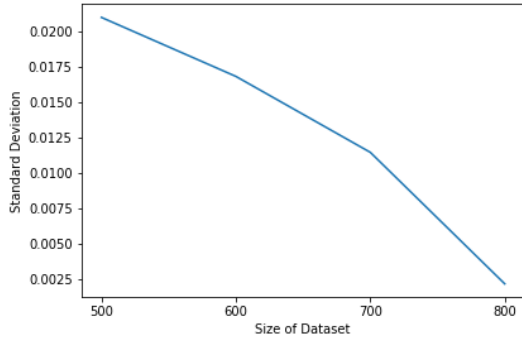
Figure 8: Mean of the standard deviation of parameter estimates as a function of dataset size. For each dataset size we fit the model 10 times with a different random seed.

5   Results and Discussion

All data and code to reproduce these results are available on Github.[3]

5.1   Alignment and Discourse Acts

Figure 6 shows the alignment estimates over our annotated discourse acts. Due the limitations of our data we limit our preliminary research question to whether these differences are significant. We conducted pairwise t-tests for the significance of the difference between the means of our alignment estimates for each discourse act. A clear majority were significant ($p >> 0.05$), with only 6.4% ($22/342$) insignificant. We therefore answer our first research question positively.

5.2   Detecting Discourse Acts

Figure 7 shows the change in ROC AUC of our logistic regression model with alignment features as compared to the baseline. In general alignment features are useful, with the net change over all discourse acts being positive. We therefore answer our second research question in the affirmative. However, arguing has taken an unexpected step backwards that requires further explanation. It could be a result of overfitting due to the small size of our dataset.

6   Reliability

Due to the limitations of our study we asked the question: how reliable are the alignment estimates presented here? We expect noise to come from three sources: (1) the small size of our dataset; (2) using a non-deterministic optimization algorithm;

(3) only having one annotator. We are unable to address (3) in the present work. However we investigated (1) and (2) by fitting our model 10 times with different random seeds for different dataset sizes (500, 600, 700, and 800 data points) and calculating the standard deviation in the estimated parameter means across the 10 runs. The results are given in Figure 8. We can see that by 800 data points the mean of the standard deviation has reduced significantly to around 0.002. Thus in the aggregate the parameters estimates appear to be converging already - although parameters with few data points still show larger variance. We clearly need more data for lower frequency discourse acts.

7   Conclusion and Future Work

We have reported what are likely to be robust results showing significant difference among alignment effects over argumentative discourse acts in a below the line comments genre. Comment level alignment features were shown to be useful for detecting argumentative discourse acts in the aggregate. Our study is limited by a small dataset, which is particularly felt for low frequency discourse acts, and an annotation scheme lacking multiple annotators. Therefore our immediate future work includes expanding our dataset and acquiring multiple annotations. We also plan to make our investigations more robust by including a cross-linguistic comparison with Mandarin data.

Although these results are not robust enough to draw more interesting conclusions about the observed patterns, we make one suggestive observation. Alignment appears higher for discourse acts that involve arguing. *Non-argumentative*, *Agreeing*, and *Pure Questioning* show no alignment effects. In general, *Arguing* and *Disagreeing* increase alignment. There is support in the previous literature for a view of alignment as modulated by engagement (Niederhoffer and Pennebaker, 2002). Our genre can be characterized as a clash of opinions. If engaging in debate is modulating alignment it would not be surprising if alignment effects were higher over argumentative discourse acts. We leave a thorough treatment of this question to future work.

We note that our agreement and disagreement estimates are at odds with previous work on body and head movement accommodation that showed alignment *decrease* with disagreement (Paxton and Dale, 2013; Duran and Fusaroli, 2017). There

---

are some considerations that may account for this discrepancy. Previous work (Doyle and Frank, 2016) showed that alignment was less pronounced in telephone than online textual conversation (Twitter). It was hypothesized that in the textual genre there is time to review the original message when composing a reply. There may also be time to reflect and choose a communication strategy. In face-to-face argumentation, on the other hand, one is forced to react in the moment, with far less time to prepare a considered response. Our tentative results appear to support a view alignment as modulated by *communication strategy* (Fusaroli et al., 2012).

We also need to apply our methods to existing datasets for comparison. In particular the US 2016 Election Reddit corpus (Visser et al., 2019) is already annotated with IAT discourse acts. The IAC should also be used to further investigate the relationship between alignment and disagreement, particularly as our finding appears to contradict previous results.

Our methods, particularly the calculation of local alignment in replying comments, can be sharpened, especially as the volume of data grows. We also note that in our dataset repliers often directly quote large portions of text in the original message. This may skew alignment calculations in these instances. We will apply a preprocessing step in future to control for this. Another peculiar feature of our genre is that comments are often directed to the broader audience. IAC is annotated with this aspect, and it will be important to investigate how this affects alignment. It may be worthwhile investigating methods that consider a broader context than the immediate message and reply. We also need to consider alignment over words as well as categories, particular as previous research showed alignment over words to be a more primary phenomenon (Doyle and Frank, 2016).

Other phenomenon have been proposed to modulate alignment in argumentation. It has been suggested that arguing a minority position may be accompanied by an increased need for persuasiveness (Pennebaker et al., 2003) (and therefore an increased usage of "causation" words). Argumentation schemes may also prove to modulate alignment. An argument from authority, for example as an eyewitness, could require a communicative strategy that sounds authoritative - having the power of knowledge. Previous results showed that power does not align but is aligned to. That would lead to the hypothesis that such an argument scheme should be correlated with a smaller or negative alignment effect. Modeling argument schemes directly may therefore help to improve the accuracy of argumentative alignment estimates.

Acknowledgements

References

Rob Abbott, Brian Ecker, Pranav Anand, and Marilyn A. Walker. 2016. Internet argument corpus 2.0: An sql schema for dialogic social media and the corpora to go with it. In *LREC*.

Frances R. Bilous and Robert M. Krauss. 1988. Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads. *Language & Communication*, 8(3):183 – 194. Special Issue Communicative Accomodation: Recent Developments.

Kasia Budzynska, Mathilde Janier, Chris Reed, and Patrick Saint-Dizier. 2016. Theoretical foundations for illocutionary structure parsing. *Journal of Argumentation and Computation*, 7(1):91–108. Thanks to IOS PRESS editor. The definitive version is available at http://www.iospress.nl/journal/argument-computation/. This papers appears in Volume 7 of Journal of Argumentation and Computation ISSN 1946-2166 The original PDF is available at: http://content.iospress.com/articles/argument-and-computation/aac005.

Katarzyna Budzynska and Chris Reed. 2011. Speech acts of argumentation: Inference anchors and peripheral cues in dialogue. In *Computational Models of Natural Argument*.

Bob Carpenter, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. Stan: A probabilistic programming language. *Journal of Statistical Software, Articles*, 76(1):1–32.

Robin Cohen. 1987. Analyzing the structure of argumentative discourse. *Computational Linguistics*, 13:11–24.

Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang, and Jon M. Kleinberg. 2011. Echoes of power: Language effects and power differences in social interaction. *CoRR*, abs/1112.3670.

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, Berlin, Germany. Association for Computational Linguistics.

Gabriel Doyle, Amir Goldberg, Sameer Srivastava, and Michael Frank. 2017. Alignment at work: Using language to distinguish the internalization and self-regulation components of cultural fit in organizations. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 603–612, Vancouver, Canada. Association for Computational Linguistics.

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, pages 637–648, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

Nicholas D. Duran and Riccardo Fusaroli. 2017. Conversing with a devils advocate: Interpersonal coordination in deception and disagreement. *PLOS ONE*, 12(6):1–25.

Riccardo Fusaroli, Bahador Bahrami, Karsten Olsen, Andreas Roepstorff, Geraint Rees, Chris Frith, and Kristian Tyln. 2012. Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological Science*, 23(8):931–939. PMID: 22810169.

Howard Giles, Nikolas Coupland, and Justine Coupland. 1991. *Accommodation theory: Communication, context, and consequence*, Studies in Emotion and Social Interaction, page 168. Cambridge University Press.

Augusto Gnisci. 2005. Sequential strategies of accommodation: a new method in courtroom. *The British journal of social psychology*, 44 Pt 4:621–43.

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(1):39–44. PMID: 21149854.

Ewa Kacewicz, James W. Pennebaker, Matthew Davis, Moongee Jeon, and Arthur C. Graesser. 2014. Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, 33(2):125–143.

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

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

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

Alexandra Paxton and Rick Dale. 2013. Argument disrupts interpersonal alignment. *Quarterly journal of experimental psychology (2006)*, 66:2092–102.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

James Pennebaker, Ryan Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of liwc2015.

James Pennebaker and Laura King. 2000. Linguistic styles: Language use as an individual difference. *Journal of personality and social psychology*, 77:1296–312.

James W. Pennebaker, Matthias R. Mehl, and Kate G. Niederhoffer. 2003. Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54(1):547–577. PMID: 12185209.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

Sara Rosenthal and Kathy McKeown. 2015. I couldn't agree more: The role of conversational structure in agreement and disagreement detection in online discussions. In *SIGDIAL Conference*.

Hagyeong Shin and Gabriel Doyle. 2018. Alignment, acceptance, and rejection of group identities in online political discourse. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 1–8, New Orleans, Louisiana, USA. Association for Computational Linguistics.

Sameer B. Srivastava, Amir Goldberg, V. Govind Manian, and Christopher Potts. 2018. Enculturation trajectories: Language, cultural adaptation, and individual outcomes in organizations. *Management Science*, 64(3):1348–1364.

Manfred Stede. 2013. From argument diagrams to argumentation mining in texts:. *International Journal of Cognitive Informatics and Natural Intelligence*, 7:1–31.

Jacky Visser, Barbara Konat, Rory Duthie, Marcin Koszowy, Katarzyna Budzynska, and Chris Reed. 2019. Argumentation in the 2016 us presidential elections: annotated corpora of television debates and social media reaction. *Language Resources and Evaluation*.

Michael Willemyns, Cynthia Gallois, Victor J. Callan, and Jeffery Pittam. 1997. Accent accommodation in the job interview: Impact of interviewer accent and gender. *Journal of Language and Social Psychology*, 16(1):3–22.

Daniel Yurovsky, Gabriel Doyle, and Michael C. Frank. 2016. Linguistic input is tuned to children's developmental level. In *CogSci*.