Linguistic Style Accommodation in Disagreements

Elise van der Pol, **Sharon Gieske** and **Raquel Fernández**
Institute for Logic, Language and Computation
University of Amsterdam
{elisevanderpol|sharongieske}@gmail.com, raquel.fernandez@uva.nl

Abstract

We investigate style accommodation in online discussions, in particular its interplay with content agreement and disagreement. Using a new model for measuring style accommodation, we find that speakers coordinate on style more noticeably if they disagree than if they agree, especially if they want to establish rapport and possibly persuade their interlocutors.

1 Introduction

In interactive communication, speakers tend to adapt their linguistic behaviour to one another at several levels, including pitch, speech rate, and the words and constructions they use. This phenomenon has been studied from different perspectives, most notably cognitive psychology and sociology. For instance, the Interactive Alignment Model (Pickering and Garrod, 2004) claims that priming mechanisms, which are an inherent feature of humans' cognitive architecture, lead to interpersonal coordination during dialogue. In contrast, Communication Accommodation Theory (Shepard et al., 2001) focuses on the external factors influencing linguistic accommodation (e.g., the wish to build rapport) and argues that converging on linguistic patterns reduces the social distance between interlocutors, which results in speakers being viewed more favourably.

Within the latter sociolinguistic view, a common methodology used to study linguistic accommodation is to focus on stylistic accommodation, as reflected in the use of function words, such as pronouns, quantifiers, and articles (Chung and Pennebaker, 2007). Previous research has shown that matching of function words signals relative social status between speakers (Niederhoffer and Pennebaker, 2002) and can be used to predict re-

lationship initiation and stability in speed dating conversations (Ireland et al., 2011). Furthermore, Danescu-Niculescu-Mizil et al. (2012) found that speakers adapt their linguistic style more when they talk to interlocutors who have higher social status and Noble and Fernández (2015) showed that this is also the case for interlocutors with a more central position in a social network.

In this paper, we investigate style accommodation in online discussions. Rather than looking into status- or network-based notions of power differences, we capitalise on the argumentative character of such discussions to study how argumentative aspects such as agreement and disagreement relate to style accommodation. In particular, we focus on the interplay between alignment of beliefs—interlocutors' (dis)agreement on *what* is said—and alignment of linguistic style—interlocutors' coordination or lack thereof on *how* content is expressed. Our aim is to investigate the following hypotheses:

H1: Speakers accommodate their linguistic style more to that of their addressees' if they agree with them on content than if they disagree.

H2: Speakers who disagree on content coordinate their linguistic style more towards addressees they want to persuade than towards those they want to distance themselves from.

Given evidence for the relationship between affiliation and mimicry (Lakin and Chartrand, 2003; Scissors et al., 2008), H1 seems a sensible conjecture. Hypothesis H2 is grounded on the assumption that individuals who disagree with their interlocutors may want to persuade them to change their mind. This creates a certain power difference, with the persuader being in a more dependent position. As shown by Danescu-Niculescu-Mizil et al. (2012), such dependence can lead to increased style matching.

2 Data

For our investigation we use the Internet Argument Corpus (IAC) (Walker et al., 2012), which contains a collection of online discussions scraped from internet fora. About 10,000 Quote-Response (Q-R) pairs have been annotated with scalar judgments over a multitude of dimensions, including level of agreement/disagreement (scale 5 to –5). Although the corpus does not include an annotation that directly indicates level of persuasiveness, we approximate persuasion by making use of two additional annotated dimensions: nice/nastiness (scale 5 to –5) and sarcasm (scale 1 to –1). We assume that responses that are perceived as nicer are more likely to be persuasive than those perceived as nasty. Similarly, we take sarcastic responses as being more likely to signal a distancing attitude than a persuasion goal.

Each Q-R pair has been judged by 5 to 7 annotators on Amazon Mechanical Turk and their scores have been averaged for each dimension. Walker et al. (2012) report relatively low inter-annotator agreement (measured with Krippendorf's α): 0.62 for agreement/disagreement, 0.46 for nice/nastiness, and only 0.22 for sarcasm.[1] We therefore chose to leverage only a subset of the corpus for which there is substantial agreement on either side of the scales. For the nice/nasty and agreement/disagreement judgments, we only consider Q-R pairs with strong majorities, i.e., Q-R pairs where all judgments except at most one are either ≥ 0 or ≤ 0. For sarcasm, we only consider Q-R pairs where there is at most one neutral judgment (value 0) and at most one judgment opposite to the majority.

In addition, to be able to assess the style of individual authors, we restrict our analysis to Q-R pairs with response authors who contribute responding posts in at least 10 different Q-R pairs. The resulting dataset after applying all these constraints contains a total of 5,004 Q-R pairs, 14% of which correspond to agreeing responses, 65% to disagreeing responses, and 21% to neutral responses. This mirrors the distribution in the full, unfiltered corpus: 13% agreeing, 67% disagreeing, and 20% neutral responses.

[1]According to Walker et al. (2012), these α scores were computed using an ordinal scale (except for sarcasm) on a dataset comprising both the set of Q-R pairs we take as starting point here and data from an additional experiment referred to as P123 by the authors. See their paper for details.

3 Measuring Linguistic Accommodation

We measure linguistic style accommodation with respect to 8 different functional markers (personal pronouns, impersonal pronouns, articles, prepositions, quantifiers, adverbs, conjunctions, and auxiliary verbs) using the lists made available by Noble and Fernández (2015).[2] Our starting point is the linguistic coordination measure proposed by Danescu-Niculescu-Mizil et al. (2012), which uses a subtractive conditional probability to capture the increase in the probability of using a marker given that it has been used by the previous conversation participant. In our notation:

$$C^m = p(R_i^m | Q_j^m) - p(R_i^m) \qquad [1]$$

Here $p(R_i^m | Q_j^m)$ refers to the probability that a response R by author i contains marker m given that the quoted post by j also contains m. How much coordination C there is in i's responses to j corresponds to the difference between this conditional probability and the prior probability $p(R_i^m)$ for author i, i.e., the probability that any response by i contains a linguistic marker of category m.

Given the sparsity of data in online discussion fora with regards to repeated interactions between the same individuals i and j, we compute a score for each Q-R pair (rather than for the set of Q-R pairs between specific authors i and j). Therefore, the conditional probability in Equation [1] corresponds to a variable that takes value 1 if both Q and R contain m and 0 if only Q does (and is undefined if Q does not contain m). The prior again corresponds to the proportion of responses by the author of R that exhibit m in the entire dataset.

A problem with this measure (both in the original formulation by Danescu-Niculescu-Mizil et al. and our own with a boolean term) is that it does not account for utterance length: clearly, a longer response has more chances to contain a marker m than a shorter response. Indeed length has been observed to be an important confounding factor in the computation of stylistic coordination (Gao et al., 2015). We therefore proposed an extension of the original measure to account for both aspects independently: the *presence* of a marker in a post (1 vs. 0) and its *frequency* given the post length.

In our model, alignment between Q and R and the prior for the author of R with respect to

[2]These lists of markers are based on Linguistic Inquiry and Word Count (LIWC) by Pennebaker et al. (2007).

marker class m correspond to feature vectors $\vec{a}$ and $\vec{b}$, respectively, with a first feature indicating marker presence and a second feature accounting for marker frequency. Thus, for a given Q-R pair:

a_1: presence of m in R given that Q contains m
a_2: proportion of words in R that are m

Similarly, for a given author i, the prior includes the following features:

b_1: proportion of responses R by i containing m
b_2: proportion of words by i that are m

After rescaling all features to range $[0, 1]$, $\vec{a}$ and $\vec{b}$ are scalarized by taking the dot product with a so-called weight vector $\vec{w}$, which determines the importance of each feature (*presence* vs. *frequency*). This *linear scalarization* is a standard technique in multi-objective optimization (Roijers et al., 2013). To determine the SA_m score of a given Q-R pair for a marker class m, as in the original measure we finally take the difference between the alignment observed in the Q-R pair and the prior encoding the linguistic style of the responding author:

$$SA_m = (\vec{a} \cdot \vec{w}) - (\vec{b} \cdot \vec{w}) \qquad [2]$$

An advantage of this measure is that it allows us to explore the effects of using different weights for different features, in our case *presence* vs. *frequency*, but potentially other features (such as syntactic alignment) as well. In the current setting, if $w_2 = 0$, we obtain the original measure where only the presence of a marker is recorded, without taking into account frequency and hence post length. In contrast, if $w_1 = 0$, only relative marker frequency is considered and no importance is given to the mere presence of a marker in a post. If the two weights are above zero, both features are taken into account.

4 Analysis and Results

For each Q-R pair in our dataset, we compute SA_m for each marker m, as well as the average style accommodation over all markers, which we refer to simply as SA. To test the hypotheses put forward in the Introduction, we retrieve clearly agreeing Q-R pairs (agreement annotation > 1, $N = 468$) and clearly disagreeing Q-R pairs (agreement annotation < -1, $N = 2519$). All our analyses are performed on these subsets.

According to hypothesis H1, more style accommodation is expected to be present in agreeing responses. We find a significant difference in SA_m
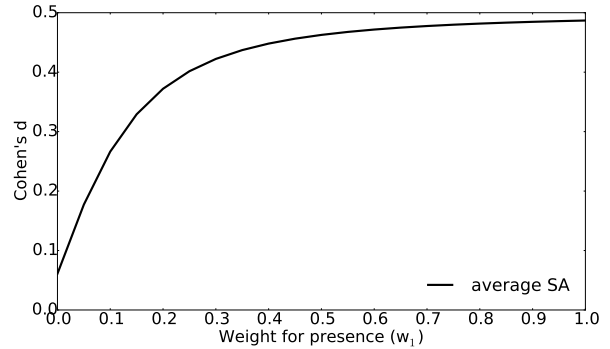

Figure 1: Effect size (Cohen's d) when comparing SA in agreeing and disagreeing Q-R pairs with different feature weights; w_1 (*presence*) in the x-axis.

for all markers between agreeing and disagreeing Q-R pairs (Welch two sample t-test, $p < 0.001$; effect size Cohen's d between .22 and .37 for all markers). Contrary to hypothesis H1, however, in all cases the level of style accommodation is higher in *disagreeing* responses than in agreeing ones. Example (1) in Table 1 shows a typical Q-R pair with high content agreement but low SA. As illustrated by this example, strongly agreeing responses often consist of short explicit expressions of agreement, with less potential for stylistic alignment. In contrast, disagreeing responses tend to be longer (as already observed by conversational analysts such as Pomerantz (1984)) and have therefore more chances to include stylistic markers matching the quoted post.

Indeed, although across the board disagreeing responses exhibit more SA, the statistical significance of this difference decreases as we lower the weight of the *presence* features (and thus give more importance to frequency and post lenght). Figure 1 shows the evolution of the effect size (Cohen's d) with different values for w_1. When only frequency is taken into account ($w_1 = 0$), the effect size is very low. However, as soon as w_1 receives some weight (from $w_1 = 0.1$ onwards), a more significant difference can be observed for disagreeing Q-R pairs (Welch two sample t-test, $p < 0.001, d > 0.2$).[3]

We now concentrate on disagreeing Q-R pairs to investigate our second hypothesis. According to H2, disagreeing responses with a persuasive

---

[3] As suggested by one anonymous reviewer, we also performed our analysis on a balanced dataset constructed by under-sampling the category of disagreeing Q-R pairs. Our findings also hold in this balanced setting.
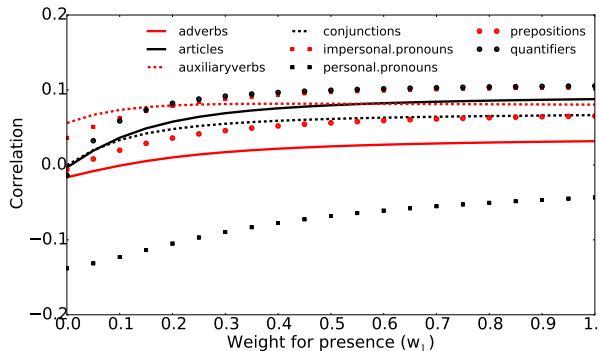
Figure 2: Correlation between SA and nice/nastiness annotations in disagreeing Q-R pairs.

Figure 3: Effect size (Cohen's d) when comparing SA in disagreeing nice vs. nasty Q-R pairs with different feature weights; w_1 in the x-axis.

(1) Q: Micheal Moore tends to manipulate people, just in a different way than the President or the media does. . . not with fear, with knowledge and anger. **R:** Well said. I agree 100%. *agreement*= 5, *nice/nasty*= 5, $SA_{avg} = -.39$ $(\vec{w} = [0.5, 0.5])$
(2) Q: And the problem is, if one of these assumption is proven incorrect, then the whole theory collapses. **R:** And one of these assumption has not been proven incorrect. *agreement*= –2, *nice/nasty*= 5 , $SA_{quant} = .31$ $(\vec{w} = [0.5, 0.5])$
(3) Q: But he does have a point. . . **R:** I see. Then you have none? *agreement*= –1, *nice/nasty*= –2, $SA_{pers.pro}=.24$ $(\vec{w} = [0.5, 0.5])$

Table 1: Example Q-R pairs.

aim will show higher style accommodation. As mentioned earlier, we use the nice/nastiness and sarcasm annotations as a proxy for persuasion or lack thereof. We observe a tendency towards positive correlation between style accommodation and niceness: i.e., responses with higher SA tend to be perceived as nicer. Nevertheless, the correlations observed, although often significant ($p < 0.05$), are extremely weak (Pearson's $r < 0.1$). Interestingly, however, in this case there is one type of marker that stands out: accommodation on personal pronouns is *negatively* correlated with level of niceness. This can be observed in Figure 2, which plots SA for all markers separately for different feature weighting schemes. As can be seen, the negative correlation for personal pronouns is stronger the more weight we give to marker frequency (lower values of w_1 in the plot). This correlation is significant ($p < 0.05$) for all values of w_2 higher than 0.1.

We next discard neutral values on the nice/nastiness dimension and focus on Q-R pairs that have clearly been annotated as nice (score > 1) or nasty (score < -1). We find significant differences for four marker types: auxiliary verbs, quantifiers, impersonal and personal pronouns. Not surprisingly, given the correlations observed above, the three former markers show more SA in nice disagreeing responses, while SA with respect to personal pronouns is higher in nasty responses. Examples (2) and (3) in Table 1 illustrate this. Figure 3 shows the effect size of these differences (Cohen's d) for these four marker types, for different feature weight values. As clearly seen in the plot, personal pronouns also contrast with the other markers on their behaviour with different weighting schemes. The
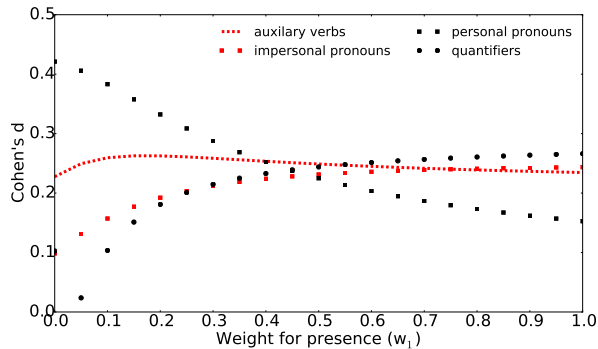
higher accommodation on personal pronouns (in nasty responses) is much more pronounced when marker frequency receives a high weight.

Finally, regarding sarcasm, we observe a tendency for all markers to exhibit lower levels of style accommodation in sarcastic disagreeing responses. This tendency is statistically significant for three marker types: auxiliary verbs, quantifiers, and impersonal pronouns (Welch two sample t-test, $p < 0.05$ for $w_1 > 0.25$). Accommodation on personal pronouns does not appear to be related to sarcasm. We remark, however, that these results need to be taken with care since only 3% of all Q-R pairs in the dataset (5% in disagreeing pairs) are reliably annotated as sarcastic.

5   Conclusions

We have investigated style accommodation in online discussions by means of a new model that takes into account the *presence* of a marker in both quoted text and response and the relative *frequency* of that marker given the length of a post.

123

Contrary to our first hypothesis, we found more accommodation in disagreeing responses than in agreeing ones. Thus, if speakers fully align on content, there seems to be less need to also align on style; in contrast, when there is a content disagreement, speakers may want to maintain rapport by exhibiting style accommodation. In support of our second hypothesis, we observed more accommodation in disagreeing responses that were perceived as nice by annotators. In a discussion, such responses are presumably more persuasive than those perceived as nasty or sarcastic, where style accommodation was lower.

We found pronounced differences for personal pronouns: in the current dataset, accommodation on personal pronouns signals distancing (nasty perception). The fact that personal pronouns stand out confirms previous findings showing that this marker class can be a particularly powerful indicator of social dynamics (Pennebaker, 2011).

Our analysis has shown that the relative weight given to *presence* and *frequency* features can have a substantial impact on the results obtained. We hope that the model put forward will help to further understand confounding factors in the computation of style accommodation. We leave a thorough investigation of these issues to future work.

Acknowledgements

References

Cindy Chung and James W. Pennebaker. 2007. The psychological functions of function words. *Social Communication*, pages 343–359.

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

Shuyang Gao, Greg Ver Steeg, and Aram Galstyan. 2015. Understanding confounding effects in linguistic coordination: an information-theoretic approach. *PloS One*, 10(6):e0130167.

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.

Jessica L. Lakin and Tanya L. Chartrand. 2003. Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science*, 14(4):334–339.

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 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, June. Association for Computational Linguistics.

James W. Pennebaker, Martha E. Francis, and Roger J. Booth. 2007. Linguistic Inquiry and Word Count (LIWC): A computerized text analysis program. Technical report, LIWC.net, Austin, Texas.

James W. Pennebaker. 2011. *The Secret Life of Pronouns: What*. Bloomsbury Press.

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

Anita Pomerantz. 1984. Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shaped. In *Structures of Social Action*. Cambridge University Press.

Diederik Marijn Roijers, Peter Vamplew, Shimon Whiteson, and Richard Dazeley. 2013. A survey of multi-objective sequential decision-making. *Journal of Artificial Intelligence Research*.

Lauren E. Scissors, Alastair J. Gill, and Darren Gergle. 2008. Linguistic mimicry and trust in text-based CMC. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pages 277–280. ACM.

Carolyn A. Shepard, Howard Giles, and Beth A. Le Poire. 2001. Communication accommodation theory. In W. P. Robinson and H. Giles, editors, *The new Handbook of Language and Social Psychology*, pages 33–56. John Wiley & Sons Ltd.

Marilyn A Walker, Jean E Fox Tree, Pranav Anand, Rob Abbott, and Joseph King. 2012. A corpus for research on deliberation and debate. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC)*, pages 812–817.