INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

A convex model for linguistic influence in group conversations

Kan Kawabata[1], *Visar Berisha*[1,2], *Anna Scaglione*[1], *Amy LaCross*[2]

[1]School of Electrical, Computer and Energy Engineering
[2]Speech and Hearing Science Department
Arizona State University, Tempe, Arizona, USA

kkawabat@asu.edu, visar@asu.edu, Anna.Scaglione@asu.edu, Amy.Lacross@asu.edu

Abstract

Conversational partners can influence each other's speaking patterns. In this paper, we aim to develop a computational model that infers influence levels directly from language samples. We propose a new approach to modeling linguistic influence in conversations based on a well-accepted model of social influence. Very generally, this approach assumes that an individual's language model can be expressed as a convex combination of language models from individuals with whom that person interacts. We propose an optimization criterion to estimate the pairwise influence between conversational partners directly from speech and language data. We evaluate the model on three different corpora: (1) a synthetic corpus where the language influence is experimentally set; (2) a corpus that tracks a child's interaction with her family during the early stages of language development; (3) a corpus of Supreme Court cases analyzing interactions between judges and attorneys.

Index Terms: linguistic influence, language accommodation, communication accommodation, convex modeling, social influence, De Groot's model

1. Introduction

Communication accommodation refers to the phenomenon that conversational partners converge to one another in communicative behavior [1]. Synchrony between partners is observed in word choice, acoustic parameters (e.g. pitch, energy), semantic patterns, etc [1, 2, 3]. This model has been confirmed by many controlled laboratory studies and a few social network studies analyzing evolving linguistic trends [1, 2, 3]; however to date limited work has been done on computational models that can characterize the influence exhibited by conversing speakers. Such a model is of great interest since linguistic accommodation plays an important role in a variety of fields, including intercultural communication [4], child development [5], human-computer dialog systems [6], etc. In this paper, our aim is to develop a new method for analyzing this important phenomenon in order to gain deeper insight in how social interactions develop in different settings.

The existing body of literature strongly suggests that linguistic accommodation is a stable and common component in speech [7, 8, 9, 10, 11, 12, 13, 14, 4, 15, 16, 17, 18]. However, the bulk of this work, conducted in laboratory settings, relies upon shadowing and other repetition tasks using only auditory stimuli. Research using actual conversation tasks has been largely restricted to conversational dyads [19, 20]. Yet, one could argue that most linguistic encounters occur in multi-speaker group conversational settings, where individual speakers' varying motivations and influence create a complex, multi-variable interaction. The existing body of work modeling these interactions has been much more limited [2, 3, 21, 22, 23] and not always applied to group conversations [24]. Exemplary of this body of research is the work by Danescu-Niculescu-Mizil, et al [3], which which focuses on specific linguistic style markers (e.g. articles, auxiliary verbs, etc.) and analyzes pairwise changes in the probability of using a word belonging to one of these categories, given another speaker's previous use of the same marker. More recently, Guo et al [2], proposed a Bayesian model of linguistic influence where they learn the parameters using Markov Chain Monte Carlo (MCMC), and evaluated it on group conversations.

The emerging field of language dynamics also provides important prior art related to language change and evolution. A number of researchers have proposed meta-theoretical models of language change motivated by population genetics and by Social Impact Theory [25, 26]. However, these models often require knowledge of latent variables not typically available during analysis (e.g. difference in social status between speakers) and aim to characterize long-term changes in language and *not* what is influencing these changes.

Motivated by recent models of social influence [27], in this paper we propose a general computational model that tracks linguistic influence in conversations in order to identify a *linguistic influence matrix (LIM)* - a matrix that describes the relative influence between any pair of participants in a conversation. In contrast to the work in [3] that focuses on specific linguistic style markers, our model is based on De Groot's model of social influence [28], originally proposed to interpret the convergence of opinions and the wisdom of crowds [28, 29].

Briefly, in accordance with De Groot's model, the proposed approach for linguistic influence assumes that an individual's language model at time $t+1$ is shaped by other participants' language models in the past using a convex model. The coefficients of this model determine the level of influence that a speaker has on another speaker. As we will show, this reduces the problem of influence to one of inferring mixture coefficients in a probabilistic mixture model. To that end, we propose an optimization criterion for solving for these coefficients directly from carefully transcribed conversational data. This is in contrast to the work by Guo [2], where the resulting optimization criteria is complicated and requires MCMC methods for estimation. We evaluate our computational model on three different examples: (1) a synthetic example where the influence is pre-defined by the experimenter; (2) longitudinal conversational samples between a child and other members of the family where we study which members of the family have the greatest level of influence on the child's language; (3) a collection of transcribed Supreme Court cases where we study relative differences in influence between

http://dx.doi.org/10.21437/Interspeech.2016-498

Supreme Court justices and lawyers.

The rest of this paper is organized as follows: In the ensuing section we provide a general overview of the model and describe an optimization criterion for inferring the LIM. In section 3, we describe in detail the three experiments we use to validate the proposed model. In section 4, we end with concluding remarks and a description of future work.

2. Proposed Method

2.1. General overview of the model

De Groot's model of social influence states that group opinions can be modeled as convex combinations of individual opinions [28]. We adopt a similar assumption for evolving language models. That is, our model assumes that in a conversation evolving between M participants, a speaker's language model is a convex combination of the language models of the individuals with whom that person interacts.

Figure 1 illustrates a sample conversation timeline for a set of M users. For each speaker i ($i \in [1...M]$), the conversation is split into a set of T epochs. For the remainder of this paper, we define the epoch for speaker i as the period of time between when the speaker starts speaking until the next time he or she speaks. In Figure 1, the timeline for speaker 1 is shown; as a result, the first epoch ends and the second epoch begins when speaker 1 starts speaking. We denote the probabilistic language model of speaker i during epoch t by $L_i^{(t)}$.

In accordance to De Groot's model of influence, a speaker's language model at $(t+1)$ is influenced by the language model of the previous epoch in the conversation. For our model we assume that $L_i^{(t+1)}$ is a mixture model of the language models of other speakers during epoch t,

$$L_i^{(t+1)} = \sum_{j \in E_i^{(t)}} W_{ij} L_j^{(t)} + W_{ii} L_i, \tag{1}$$

where $E^{(t)i}$ is the subset of speakers that spoke during epoch t, W_{ij} is the influence coefficient that describes the probability that speaker i will imitate (be influenced by) the language of speaker j, L_i is the default (time-independent) language model for speaker i, and W_{ii} is the influence of this model on the language model used at time $t+1$. The value W_{ij} is the i^{th} column and j^{th} row entry of the $M \times M$ linguistic influence matrix (*LIM*), $\mathbf{W}$, that describes the pairwise influence between two speakers in a conversation. The LIM coefficients are all non-negative entries and the sum of each row is 1 (i.e. $\sum_{j=1}^{M} W_{ij} = 1$).

2.2. Learning the Language model

The model above requires an estimate of the probabilistic language model for each participant during each epoch in a conversation. These could be unigram, trigram, or polygram models inferred from the text data. For shorter texts these could also be probabilistic features learned using semantic processing and linguistic style markers [30]. Such feature models are especially useful for corpora with sparse data (e.g. twitter conversations) [3]. For larger corpora, these models could by based on deep networks [31].

For the remainder of this work, for simplicity, we use a simple unigram model of $L_i^{(t)}$. We assume the domain of the unigram model is the vocabulary $\mathcal{V} = \{w_1, \ldots, w_V\}$ which in the following will be addressed simply with the numbers

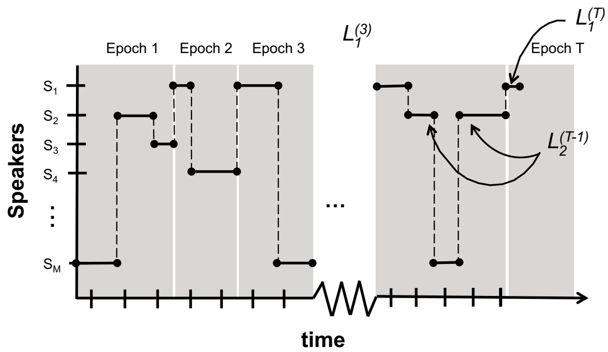

Figure 1: A conversational timeline depicting an interaction between M speakers

$v = 1, \ldots, V$. Given a specific corpus, we can estimate a probabilistic language model for each speaker at each time t, $L_i^{(t)}(v) = P(w_v|i;t)$, for each word $w_v \in \mathcal{V}$, and speaker $i = 1, \ldots, M$. Given the text from epoch t for speaker i, we can estimate $L^{(t)}$ as

$$\widehat{L}_i^{(t)}(v) = \frac{F_i^{(t)}(v)}{\sum_v F_i^{(t)}(v)}, \tag{2}$$

where $\widehat{L}_i^{(t)}$ is the estimate of the true language model $L_i^{(t)}$ and $F_i^{(t)}(v)$ is the frequency of the word w_v for speaker i during epoch t.

In eqn. (1), we make a distinction between the time-varying language models of the participants in a conversation and the default language model for speaker i. For the experiments in this paper, we estimate speaker i's default model using the complete corpus. That is, we estimate the normalized frequency of each word for speaker i for the duration of the conversation,

$$\widehat{L}_i(v) = \frac{\sum_{t=1}^{T} F_i^{(t)}(v)}{\sum_{t=1}^{T} \sum_v F_i^{(t)}(v)}. \tag{3}$$

It is important to note that it isn't necessary to estimate the default model from the same corpus from which we estimate the LIM coefficients. In fact, it may be advantageous to estimate default models using many large and varying language samples from the person in different settings.

2.3. Solving for the influence coefficients

After estimating the language model for each participant during each epoch, we can construct an optimization criterion for estimating the individual entries of the LIM for a particular speaker i, $W_{i1} \ldots W_{iM}$, using the model from eqn. (1). Specifically, we assume that the estimated unigram language model for speaker i at time $t+1$ can be expressed as a mixture of unigram language models from other speakers and speaker i's default language model. The magnitude of the mixture coefficients serve as an estimate of the influence that various conversational partners have on speaker i. This influence can be compared against the user's preference for his or her default language model (W_{ii}).

The optimization criterion to solve for the LIM entries is

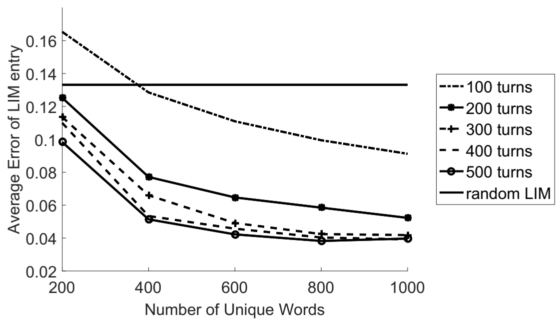

Figure 2: Average RMSE LIM entry error for different number of speaking turns and vocabulary size.

given by:

$$\begin{aligned} \underset{W_{i1}\ldots W_{iM}}{\text{minimize}} \quad & \sum_{t=1}^{T-1} \Big|\Big| \sum_{j \in E_i^{(t)}} W_{ij}\widehat{L}_j^{(t)} + W_{ii}\widehat{L}_i - \widehat{L}_i^{(t+1)} \Big|\Big|_2 \\ \text{subject to} \quad & \sum_{j=1}^{M} W_{ij} = 1, \\ & W_{ij} \geq 0,\ i = 1, \ldots, M. \end{aligned} \tag{4}$$

where $E_i^{(t)}$ represents the subset of speakers that spoke during speaker i's t^{th} epoch. The convex criterion finds the best fit parameters to eqn. (1) subject to the constraints that the weights sum to one and are non-negative. The error in the fitting routine is measured by the ℓ_2 norm. We do this for simplicity here; however, since we are comparing two probability mass functions, any divergence measure between distributions can be used. For example, we can replace the ℓ_2 norm of the error with the Kullback-Leibler divergence or the data-driven D_p divergence [32, 33] between the model at $t+1$ and the mixture model from eqn. (1). Since we use the ℓ_2 norm in this paper, the resulting optimization problem is convex and we can use the CVX package to solve for the LIM coefficients [34].

The solution to (4) represents a single row of the LIM (the row corresponding to speaker i); thus to solve for all values of LIM the optimization in criterion (4) must be solved individually for each speaker. By analyzing differences between LIM coefficients, we can compare the relative levels of influence on speaker i's language model from other speakers, and how these compare to speaker i's preference for his or her default language model.

3. Experiments and Results

3.1. Experiment 1: Academic Example

Experiment design: In order to show validity of our method we simulated a synthetic example of a conversation between 6 speakers based on a randomly generated 6×6 LIM. Each speaker is initialized with a default unigram model from a domain containing V words. The probabilities for the unigram model are drawn from a Poisson distribution, with 5% of the entries randomly set to zero to ensure diversity in the vocabulary of each speaker.

We run the simulation for T conversational turns where, for each turn, a speaker will generate 100-200 words[1] based on a linear combination of language models as in eqn. (1). That is, the mixture model in eqn. (1) is used to draw the words for speaker i's t^{th} turn, where the mixture coefficients are random, but known to the experimenter. The purpose of the experiment is to infer the LIM coefficients using the proposed algorithm from only the text data. We ran the experiment with varying V ($V = [200, 400, 600, 800, 1000]$) and T ($T = [100, 200, 300, 400, 500]$) with 30 Monte-Carlo trials per combination.

[1] Based on statistics found in the court transcripts in Experiment 3

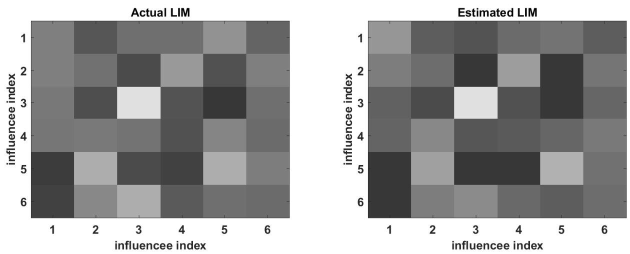

Figure 3: Sample actual and estimated LIM values for ($V = 1000, T = 200$)

We can determine how well the algorithm infers the influence dynamics of each speaker by analyzing the root mean squared error between the known LIM and the LIM estimated by our algorithm,

$$e = \sqrt{\frac{1}{M^2}\sum_{i}^{M}\sum_{j}^{M}(\hat{W}_{ij} - W_{ij})^2}, \tag{5}$$

where $\hat{W}_{ij}$ is the estimate of W_{ij}.

Results: Figure 2 shows the average error found for different values of parameters described in the setup. The figure shows that the accuracy of the algorithm improves for larger vocabulary size and for longer conversations. The figure also shows the average error value for random LIMs compared with the true LIMs as a benchmark for the performance of our model. From this we see that only for conversation of 100 turns and 200 unique words does our model fare worst than random chance, however for the real world conversations we analyzed in this study (see next sections), their parameters fall in the range that yield a significant improvement over random chance. In Figure 3 we show the estimated LIM and the actual LIM for a sample conversation with 6 speakers, 200 turns, 100-200 words per turn, and 1000 unique words. These parameters are chosen to mimic those found in the supreme court transcripts discussed in later section.

3.2. Experiment 2: Early Development Corpus

Experiment design: We evaluate the model on a language development corpus of 700 transcripts of a single child conversing with her family during early development age (approximately ages 2 - 3.5) [5]. Approximately one transcript per day was recorded and each transcript contains on average 200 lines of dialog between the child and family members. In this experiment, for each transcript, the linguistic influence of the family members on the child's vocabulary is estimated using our model. By analyzing how the LIM coefficients associated with each family member's influence we can see how influence of each member changes over time.

Results: For each transcript that contains an interaction between the child and at least one other member of the family,

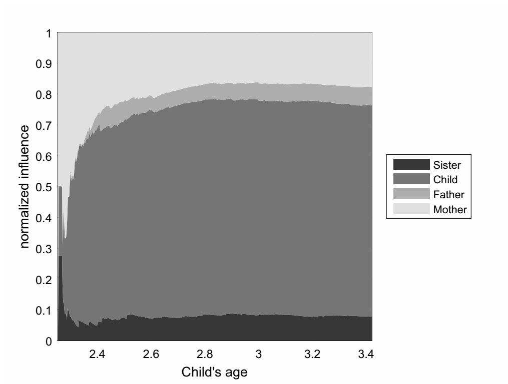

Figure 4: Running average influence of family member

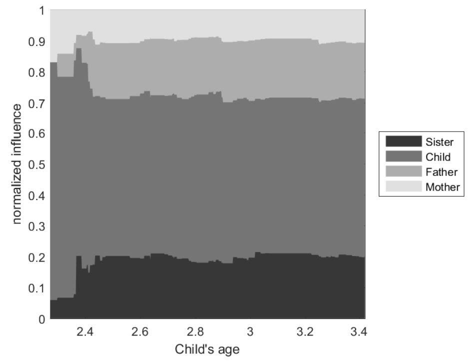

Figure 5: Running average influence of family member (all members present)

we calculate the influence of the family member on the child's vocabulary using the algorithm proposed in this paper. In Figure 4 we show the running average estimate of the influence coefficients for the duration of the period. From the figure, we can interpret that by age 3 the child's default language model (dark gray) contributes to two-third of her word choices, followed second by the mother's language model (light gray), followed by the sister (black) and the father (medium gray). The time-varying nature of this influence is interesting to note; initially the child's speech seemed to be dictated mostly by her sister's and mother's language models; however as time goes on she seems to develop her own individual language model distinct from her family. A number of studies in child development have shown similar patterns: 1) children are more likely to repeat words during early stages of language development and 2) the mother has the largest influence on language development in children [35].

By analyzing the data it is clear that the mother has the largest number of interactions with the child. As a result, it's unclear from the data whether her influence is driven by the fact that the child and the mother interact more or whether the child is inherently more influenced by the mother. To answer this question, we only analyze the subset of transcripts where all four family members are present. Unfortunately, this reduces the data from 700 transcripts to 90 transcripts; therefore the resulting analysis is rather coarse. In Figure 5 we see that the influence levels even out and, in fact, the mother's contribution is reduced significantly. This result suggests that linguistic influence is, to an extent, dependent on length of interaction between communicating parties.

3.3. Experiment 3: Supreme Court Cases

Experiment design: In this experiment we analyzed a corpus of 94 U.S. Supreme Court oral argument transcripts for cases between 2000 - 2001. The documents were found on the official U.S. Supreme Court website [36]. Due to variability in the transcription methodology, we simplified the transcripts by grouping each speaker into one of three categories: judge, defendant, or prosecutor. The 3×3 LIM is calculated for each transcript; the total influence of each group is based on the sum of the columns of the LIM. This allows us to estimate the average influence of one of the parties on the remaining two parties for all 94 transcripts.

Results: Figure 6 shows the average influence of each of the three groups on the remaining group (with $1-\sigma$ error bars). The

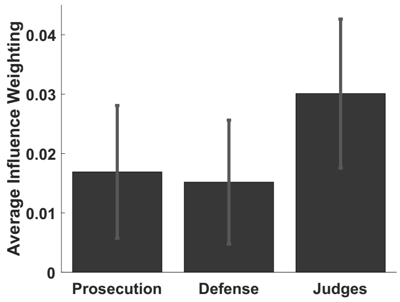

Figure 6: Influence in Supreme Court (2000-2001)

analysis reveals that Supreme Court justices have the largest influence on the remaining parties, followed by the defendant and prosecution. A two-samples t-tests reveals a statistically significant difference between the average influence of judges ($\mu = 0.333$, $\sigma = 0.060$) and the average influence of prosecutors ($\mu = 0.224$, $\sigma = 0.062$) with $p < 0.0001$. Similarly the average influence of judges was also different than the influence of defendants ($\mu = 0.240$, $\sigma = 0.022$) with $p < 0.0001$. These results are consistent with the work of Niculescu-Mizil, et al. which concluded that there is significantly more accommodation by the attorneys to the judges than vice-versa [21].

4. Conclusions

We present a new computational model to quantify linguistic influence between speakers in group conversations. The new model is based on an existing model of social influence - De Groot's convex model for opinion fusion. Based on this model we proposed an optimization criterion that is computationally simple to solve and versatile to model data from different applications. We evaluated this model on three data sets and compared the results of the model to other results from the literature.

Future work will focus on using other divergence measures for calculating the distance between language models. Furthermore, we will also explore methods for joint estimation of the influence matrix and the language model. In addition, while we focus on text data in this application, the approach can also be extended to modeling acoustic data (e.g. for modeling prosodic influence). In that case, $L_i^{(t)}$ could be a probability density function of features that measure an acoustic feature of interest (e.g. pitch, speaking rate, etc.).

5. References

[1] C. Gallois and H. Giles, "Communication accommodation theory," *The International Encyclopedia of Language and Social Interaction*, 2015.

[2] F. Guo, C. Blundell, H. Wallach, and K. Heller, "The bayesian echo chamber: Modeling social influence via linguistic accommodation," *arXiv preprint arXiv:1411.2674*, 2014.

[3] C. Danescu-Niculescu-Mizil, M. Gamon, and S. Dumais, "Mark my words!: linguistic style accommodation in social media," in *Proceedings of the 20th international conference on World wide web*. ACM, 2011, pp. 745–754.

[4] V. Barker, H. Giles, C. Hajek, H. Ota, K. Noels, T.-S. Lim, and L. Somera, "Police-civilian interaction, compliance, accommodation, and trust in an intergroup context: International data," *Journal of International and Intercultural Communication*, vol. 1, no. 2, pp. 93–112, 2008.

[5] S. R. Braunwald, "The development of because and so: Connecting language, thought, and social understanding," *Processing interclausal relationships: Studies in the production and comprehension of text*, pp. 121–137, 1997.

[6] L. Fais, "Lexical accommodation in machine-mediated interactions," in *Proceedings of the 16th conference on Computational linguistics-Volume 1*. Association for Computational Linguistics, 1996, pp. 370–375.

[7] J. Small, S. M. Chan, E. Drance, J. Globerman, W. Hulko, D. OConnor, J. Perry, L. Stern, and L. Ho, "Verbal and nonverbal indicators of quality of communication between care staff and residents in ethnoculturally and linguistically diverse long-term care settings," *Journal of cross-cultural gerontology*, vol. 30, no. 3, pp. 285–304, 2015.

[8] H.-G. Bosshardt, C. Sappok, M. Knipschild, and C. Hölscher, "Spontaneous imitation of fundamental frequency and speech rate by nonstutterers and stutterers," *Journal of psycholinguistic research*, vol. 26, no. 4, pp. 425–448, 1997.

[9] H. Giles, J. Coupland, and N. Coupland, *Contexts of accommodation: Developments in applied sociolinguistics*. Cambridge University Press, 1991.

[10] S. D. Goldinger, "Echoes of echoes? an episodic theory of lexical access." *Psychological review*, vol. 105, no. 2, p. 251, 1998.

[11] D. Bulatov, "The effect of fundamental frequency on phonetic convergence," 2009.

[12] M. Babel and D. Bulatov, "The role of fundamental frequency in phonetic accommodation," *Language and Speech*, vol. 55, no. 2, pp. 231–248, 2012.

[13] C. Gambi and M. J. Pickering, "Prediction and imitation in speech," *Frontiers in psychology*, vol. 4, 2013.

[14] M. E. Babel, "Phonetic and social selectivity in speech accommodation," Ph.D. dissertation, University of California, Berkeley, 2009.

[15] J. J. Bradac, A. Mulac, and A. House, "Lexical diversity and magnitude of convergent versus divergent style shifting-: Perceptual and evaluative consequences," *Language & Communication*, vol. 8, no. 3, pp. 213–228, 1988.

[16] S. Brouwer, H. Mitterer, and F. Huettig, "Shadowing reduced speech and alignment," *The Journal of the Acoustical Society of America*, vol. 128, no. 1, pp. EL32–EL37, 2010.

[17] N. COUPLAND, "Accommodation at work: Some phonological data and their implications," *International Journal of the Sociology of Language*, vol. 1984, no. 46, pp. 49–70, 1984.

[18] F. M.-s. Yau, "Code switching and language choice in the hong kong legislative council," *Journal of Multilingual and Multicultural Development*, vol. 18, no. 1, pp. 40–53, 1997.

[19] C.-C. Lee, M. Black, A. Katsamanis, A. C. Lammert, B. R. Baucom, A. Christensen, P. G. Georgiou, and S. S. Narayanan, "Quantification of prosodic entrainment in affective spontaneous spoken interactions of married couples." in *INTERSPEECH*, 2010, pp. 793–796.

[20] R. Levitan and J. B. Hirschberg, "Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions," 2011.

[21] C. Danescu-Niculescu-Mizil, L. Lee, B. Pang, and J. Kleinberg, "Echoes of power: Language effects and power differences in social interaction," in *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012, pp. 699–708.

[22] A. M. B. Liu, "Analysis of linguistic style accommodation in online debates," in *COLING*, 2012, pp. 1831–1846.

[23] D. M. Romero, R. I. Swaab, B. Uzzi, and A. D. Galinsky, "Mimicry is presidential linguistic style matching in presidential debates and improved polling numbers," *Personality and Social Psychology Bulletin*, vol. 41, no. 10, pp. 1311–1319, 2015.

[24] S. Gerrish and D. M. Blei, "A language-based approach to measuring scholarly impact." in *ICML*, vol. 10, 2010, pp. 375–382.

[25] L. Milroy, "Language change and social networks," 1987.

[26] D. Nettle, "Using social impact theory to simulate language change," *Lingua*, vol. 108, no. 2, pp. 95–117, 1999.

[27] A. G. Dimakis, S. Kar, J. M. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, 2010.

[28] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.

[29] F. Galton, "Vox populi (the wisdom of crowds)," *Nature*, vol. 75, pp. 450–51, 1907.

[30] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: Liwc and computerized text analysis methods," *Journal of language and social psychology*, vol. 29, no. 1, pp. 24–54, 2010.

[31] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[32] J. M. Joyce, "Kullback-leibler divergence," in *International Encyclopedia of Statistical Science*. Springer, 2011, pp. 720–722.

[33] V. Berisha, A. Wisler, A. O. Hero III, and A. Spanias, "Empirically estimable classification bounds based on a nonparametric divergence measure," *Signal Processing, IEEE Transactions on*, vol. 64, no. 3, pp. 580–591, 2016.

[34] M. Grant and S. Boyd, "Cvx: Matlab software for disciplined convex programming."

[35] C. Saint-Georges, M. Chetouani, R. Cassel, F. Apicella, A. Mahdhaoui, F. Muratori, M.-C. Laznik, and D. Cohen, "Motherese in interaction: at the cross-road of emotion and cognition?(a systematic review)," *PloS one*, vol. 8, no. 10, p. e78103, 2013.

[36] Supreme court of the united states: Argument transcript. [Online]. Available: http://www.supremecourt.gov/oral_arguments/argument_transcript.aspx