INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

A Psychology-Driven Computational Analysis of Political Interviews

Darren Cook[1], *Miri Zilka*[2], *Simon Maskell*[1], *Laurence Alison*[1]

[1]University of Liverpool, UK
[2]University of Sussex, UK

{d.cook, smaskell, alisonl}@liverpool.ac.uk, m.zilka@sussex.ac.uk

Abstract

Can an interviewer influence the cooperativeness of an interviewee? The role of an interviewer in actualising a successful interview is an active field of social psychological research. A large-scale analysis of interviews, however, typically involves time-exorbitant manual tasks and considerable human effort. Despite recent advances in computational fields, many automated methods continue to rely on manually labelled training data to establish ground-truth. This reliance obscures explainability and hinders the mobility of analysis between applications. In this work, we introduce a cross-disciplinary approach to analysing interviewer efficacy. We suggest computational success measures as a transparent, automated, and reproducible alternative for pre-labelled data. We validate these measures with a small-scale study with human-responders. To study the interviewer's influence on the interviewee we utilise features informed by social psychological theory to predict interview quality based on the interviewer's linguistic behaviour. Our psychologically informed model significantly outperforms a bag-of-words model, demonstrating the strength of a cross-disciplinary approach toward the analysis of conversational data at scale.

Index Terms: conversation analysis, automation, interviewing, communication accommodation theory

1. Introduction

Analysing interviews to deduce how an interviewer may steer an interviewee in the desired direction is not a new endeavour. Harnessing the computational power of machine learning in favour of this undertaking is a more recent development. Traditionally, analysis of conversation focused on qualitative study and required considerable manual effort, limiting the volume of analysed data. Offsetting time-exorbitant tasks to a computer presents the promise of scalability, alongside reproducibility and, debatably, greater objectivity.

Conversational outcomes have been computationally evaluated in domains such as group collaboration [1], job interviewing [2], speed-dating [3], hostage negotiation [4], and police interrogation [5]. Typically, computational models rely on manually pre-labelled data to separate the successful interactions from the non-successful ones. This reliance gives rise to potential limitations. Notably, the success measure is often not well defined, hindering the transparency of the model and as a result, its mobility between applications. In addition, domain-specific knowledge may be encoded within the labelled data, but it is not decoded into the model itself, limiting the utility of the model and any conclusion drawn with respect to the domain of origin.

To address these issues, we must develop and encourage the use of cross-disciplinary approaches. Ideally, we can capitalise on domain knowledge and expertise while exploiting the utility of computation. This paper aims to demonstrate the power of cross-disciplinary analysis by employing social psychological insight to engineer automated features and outcome measures for the benefit of interview analysis. In this paper, we demonstrate our approach on political interviews. Important in their own right, political interviews make an excellent model for conversations where the objectives of the interviewer may be in tension with the objective of the interviewee.

This paper seeks to answer two main questions: (1) can we predict, computationally, if a human will regard an interview as successful without relying on pre-labelled data? (2) Can we predict the success of an interview from the behaviour of the interviewer? The latter question originates from existing social psychological theories [6]. Specifically, communication accommodation theory [7] postulates that speakers modulate their linguistic behaviour relative to one another to meet particular objectives. Research has demonstrated that increased accommodation activity contributes to social effects such as personal attractiveness [8], and conversational fluidity [9].

Our contributions are as follows: A flexible framework for computationally analysing success in non-labeled interview transcripts, validated by a small-scale study with human participants. A set of engineered features, informed by social scientific theory, used to predict success in an interview based on the behaviour of the interviewer. We demonstrate that these domain-informed features significantly outperform a bag-of-words model. In addition, our results feed back to domain knowledge, corroborating that interviews operate in a manner consistent with less-adversarial forms of conversation.

2. Methods

2.1. Corpus pre-processing

2.1.1. Dataset

We created a corpus of N=684 political interviews from transcripts of six US cable news networks (CNN, NBC, ABC, MSNBC, CBS and Fox). We included interview segments comprising a single interviewer and a sole interviewee. We considered an interview *political* if the interviewee was a government member or had a clear affiliation with a political party. Interviews were conducted between 2013-2020 and featured 261 participants (55 interviewers and 206 interviewees). Transcripts were opportunistically-sampled from online repositories, stored as plain text files, and spot-checked to ensure accuracy. The length of interviews varied between 549 and 9102 words (M=1883.3, SD=1113.35). In total, the corpus comprised just under 1.3m words and 28,022 speech turns (M=40.97, SD=47.46).

2.1.2. Data cleaning

The python script produced to clean and standardise the transcripts is available on GitHub under an MIT licence. All non-ASCII characters and punctuation symbols were removed

http://dx.doi.org/10.21437/Interspeech.2021-2249

Table 1: *Sample of pre-processed text. NG, LC and DR refer to n-gram, lexical category, and dependency relation models, respectively.*

Model	Example
NG	"The" (unigram), "Quick" (unigram), "Brown Fox" (bigram), "Jumps Over" (bigram), "The Lazy Dog" (trigram)
LC	"The" (article), "Quick" (relativity), "Brown" (perception), "Jumps" (motion), "Over" (power), "The" (article), "Lazy" (neg_emotion)
DR	"The" (determiner), "Quick" (adjectival modifier), "Brown" (adjectival modifier), "Fox" (nominal subject), "Jumps" (ROOT), "Over" (preposition), "The" (determiner), "Lazy" (adjectival modifier), "Dog" (preposition object)

along with timestamps, annotations, and prerecorded segments. We also expanded contracted words using the dictionary provided in [10]. Transcripts were split into a sequence of speech turns. Each speech turn ended when the conversational floor was passed from one speaker to another.

2.2. Outcome metrics

Determining the quality of an interviewee response is challenging. The literature points us to completeness and truthfulness [11], and clear articulation [12]. Politicians, however, have a reputation for equivocation and evasiveness, answering less than half the questions put to them [13]. Uncooperative politicians have been shown to make superfluous comments [14], or rely on repetition of key phrases as a diversionary tactic [15]. In this work, we regard an interview successful if the interviewee answered questions fully, directly and clearly. Accordingly, we have constructed four measures of interviewee responses: specificity, diversity, relevance, and clarity. These are broadly influenced by Gricean conversational maxims of quality, quantity, relation and manner [16].

Interviewee speech turns were tokenised, part-of-speech tagged and lemmatised using the `Stanza` neural pipeline [17]. Post calculation, all success measures were normalised.

Our measure of **specificity**, inspired by research in investigative interviewing [18, 19], quantifies interviewee references to key details such as people, objects, locations, and temporal details. We automated this process using `spaCy`'s [20] Named Entity Recognizer (NER). To account for differences in interview length, we normalise unique named entity counts over the number of noun phrases uttered by the interviewee.

Clarity measures the average concreteness of interviewee speech. Concreteness is a psycholinguistic feature that refers to the degree of ambiguity of a word. We measure word concreteness scores using an established dictionary based on prior psycholinguistic research [21]. We take the average concreteness score over all interviewee words as our clarity measure.

An unwillingness to engage in conversation can be demonstrated by self repetition. Conversely, linguistic **diversity** has been linked with honesty and trustworthiness [22] [23]. For our measure of diversity, we use the global type-token ratio of interviewee speech, i.e. we divide the number of unique interviewee words by the total number of interviewee words.

Relevance reflects the extent an interviewee's response shares semantic similarity with the question they were asked. First, each non-stop-word is transformed into a vector using pre-trained `GloVe` word embeddings [24]. We then create a turn-level vector by averaging over all word embeddings within a speech turn. We calculate the similarity of each interviewee response with the preceding question via cosine similarity (the cosine of the angle between two non-zero vectors). Relevance is computed as the mean score over all question-answer pairs.

2.3. Features

We set out to predict the success of an interview from the behaviour of the interviewer. This approach was informed by communication accommodation theory, which posits that the extent speakers converge, maintain or diverge from each other linguistically correlates with their social goals [7]. Convergence (also known as mirroring, alignment or entrainment) indicates a shared understanding between speakers [25], and is associated with success in collaborative tasks [1], and increased compliance and cooperation [26]. Furthermore, higher-levels of convergence can make an individual more likeable to those they are mirroring [27]. Conversely, reticence to converge can reflect a desire to maintain personal identity [7]. Given the dynamics of an interview, we might expect divergence to play a more prominent role in the interaction; speakers in an interview occupy particular roles, whereby the interviewer asks questions that the interviewee is expected to answer. We therefore expect the interviewer to converge on certain linguistic features, relating to the topic under discussion, and diverge on others, such as interrogatives: *who, what, where, why, when*.

2.3.1. Local accommodation

We distinguish between two types of interviewer accommodation: local accommodation and global alignment. For local accommodation (LA), we follow the probabilistic framework described in [28]. This approach computes a token-level probability of the interviewer mirroring the interviewee by comparing the probability of the interviewer speaking a word after it was spoken by the interviewee and otherwise in the conversation:

$$LA^F_{(i,j)} \triangleq P(T^F_i | T^F_j, T_i \hookrightarrow T_j) - P(T^F_i | T_i \hookrightarrow T_j) \quad (1)$$

Here, the term on the left, $LA^F_{(i,j)}$ is the local accommodation of feature F by speaker i in relation to speaker j. The first term on the right is the conditional probability of speaker i, uttering F, given its previous usage by speaker j. The second term on the right is the total probability of speaker i uttering F over all replies to speaker j. We compute a score between -1 and 1 for each feature. Positive values indicate convergence, and negative values indicate divergence.

We apply equation (1) to three separate bag-of-features models: an n-gram (NG) model, where we calculate scores for unigrams, bigrams and trigrams, restricting unigrams and bigrams to 200 features each; a lexical categorization (LC) model, based on 70 categories generated by the Linguistic Inquiry Word Count (LIWC) tool [29]; and a dependency relations (DR) model where we perform dependency parsing via `spaCy` [20], and use the dependency relation tags as features. Table 1 provides an example of pre-processed text for each model.

2.3.2. Global alignment & meta features

We constructed global alignment features based on existing computational linguistic research:

Global Language Style Matching (gLSM) is a speaker-independent similarity measure of linguistic style, i.e., speakers' alignment of semantically neutral function words such as pronouns, adjectives, and adverbs [30]. We calculate the gLSM score between the two speakers as outlined in [3], and use the score as a single feature.

For the four measures defined below, we use the mean, min, and max scores as features:

Reciprocal Language Style Matching (rLSM) measures the extent one speaker is matching the linguistic style of another on a turn-by-turn level. We calculate rLSM per interviewer speech turn as defined in [31].

Turn length difference is measured in words as the difference between each interviewer turn and the preceding interviewee turn.

Semantic relatedness measures semantic similarity via cosine similarity between each interviewer turn and the preceding interviewee turn.

Branching factor difference. In a tree-structure, the branching factor is the average number of child nodes. By performing dependency parsing on each speech turn, the branching factor becomes a proxy for syntactic depth. We compute the average branching factor over all sentences in a speech turn, and calculate the absolute difference between each interviewer turn and the preceding interviewee turn.

Variables from the meta-data collected for each interview were also included as one-hot encoded categorical variables. These included: interview length (in words); host network; political orientation of each speaker; if the speakers shared a political orientation; if the speakers were of the same gender.

2.4. Supervised machine learning

To predict each outcome metric from the features described in section 2.3, we used a forty-tree random forest with default hyper-parameters. Alternative ensemble-based supervised learning algorithms including extra trees, gradient boosting, and XGBoost achieved equal performance. Models were cross validated using K-fold, with k=10. We evaluated model performance using the root mean squared error (RMSE). We employ two baselines: an estimator that repeated the training mean (B1), and a bag-of-words model based on interviewer word frequency counts (B2).

2.5. Manual validation of computational outcome metrics

We conducted a validation study on the computational outcome metrics described in section 2.2. Ten interviews were randomly selected from the corpus, ensuring a varied distribution of outcome scores. We recruited eight human raters, naive to the purposes of the study, to assess each interview. Raters scored each interview on a one-to-ten scale per outcome metric. For example, *to what extent did the interviewer express themselves in a clear manner?* Participants also provided an overall quality score based on the perceived quality of interviewee responses. We performed an intraclass correlation (ICC) analysis on each outcome metric to measure the degree of agreement amongst our human raters. We consider ICC scores above 0.8 to indicate 'good' inter-rater agreement. The normalised computational scores were then compared to normalised human ratings.

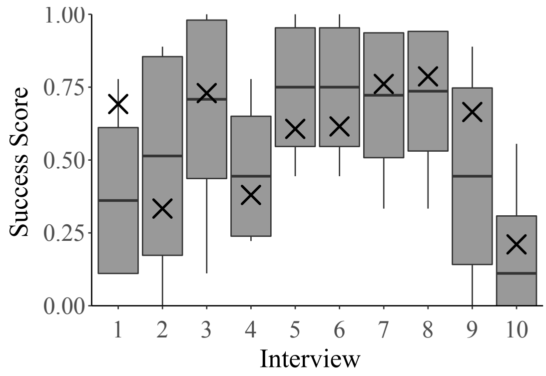

Figure 1: *Normalised distribution of human ratings of overall interview quality. The overlaid X shows the corresponding normalised computational score.*

Table 2: *Mean ($\pm SD$) RMSE scores per model iteration when predicting Clarity, Diversity, and Relevance.*

Model	Clarity	Diversity	Relevance
B1	0.144±0.016	0.158±0.016	0.142±0.011
B2	0.138±0.017	0.123±0.012	0.129±0.014
Local (All)	0.128±0.013	0.114±0.013	0.125±0.012
NG Only	0.129±0.013	0.115±0.013	0.126±0.013
LC Only	0.134±0.018	0.123±0.007	0.133±0.015
DR Only	0.134±0.015	0.121±0.008	0.137±0.014
Global	0.133±0.016	0.118±0.015	0.120±0.010
Meta	0.160±0.016	0.081±0.008	0.153±0.008
Loc. + Glo.	0.124±0.010	0.108±0.014	0.115±0.010
All	0.124±0.011	0.075±0.007	0.115±0.009

3. Results and discussion

3.1. Analysis of outcome metrics

We observed good agreement (ICC $>$.8) per outcome amongst our eight human raters. Comparing the computational scores to the normalised distribution of human ratings, we find the following percentage of computational scores that fell within one standard deviation of the mean human score: specificity = 70%, clarity = 50%, diversity = 80%, relevance = 70%.

To create an overall quality score computationally, we used a linear regression model to measure how human-raters' overall rating was weighted by their individual ratings for clarity, diversity and relevance (Specificity was omitted as our computational models did not show improvement over the baseline (B1)). The appropriate coefficients were extracted and applied to the computational scores to create an overall success score (S):

$$S = 0.23 \times Clr + 0.39 \times Div + 0.42 \times Rel \tag{2}$$

Figure 1 shows the normalised distribution of overall success scores given by human raters per interview, with the computational score marked with an X. We find a 90% agreement with the mean human score (computational scores within one SD). Based on the average variance between the mean human score and the computational score, it would require four human raters to improve on the computational predictions.

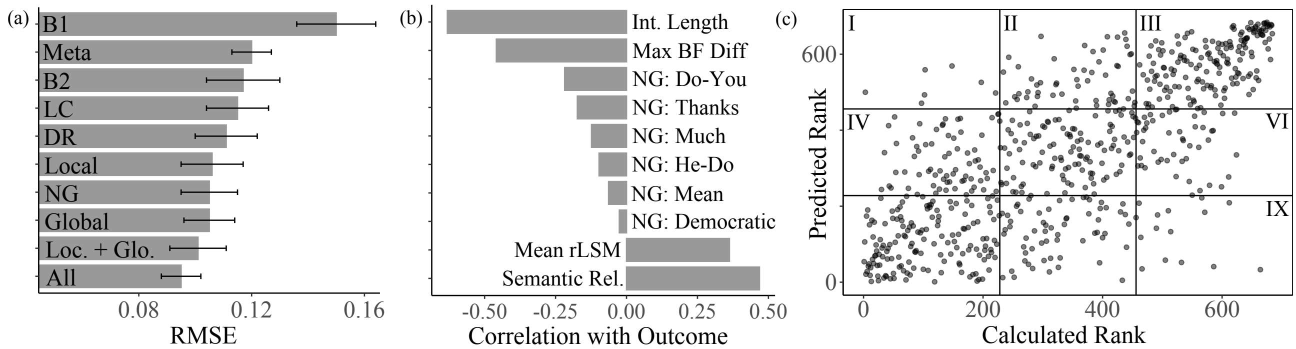

Figure 2: *2a. (left) - Prediction performance for each model iteration on the overall success measure; 2b (centre) - Correlation of the top ten features on the All Features model with the overall success score; 2c. (right) - Relationship between rank-ordered ground-truth and prediction for the Loc. + Glo. model predicting overall interview success.*

3.2. Prediction performance

Table 2 reports the RMSE for each model iteration for the clarity, diversity, and relevance metrics. The performance for overall success is shown in Figure 2 (a). One-tailed Wilcoxon signed-rank tests indicated that absolute model errors for clarity, diversity, relevance, and overall success were significantly lower than B1 and B2 ($p < .001$), but specificity was not. The combination of local and global accommodation features (Loc. + Glo. model) exceeded both baselines for clarity, diversity, relevance, and overall success. Meta features alone only exceeded the baseline when predicting diversity. The addition of the meta features to the local and global features (All model) only improved performance further for diversity and consequently for the overall success. When predicting overall success, the biggest improvement in RMSE is for the combined model (All), with a 37% improvement on B1 and 19% on B2.

In Figure 2 (b), we report the most important features for the *All* model when predicting overall success. Using permutation feature importance [32], we report the strength and direction of correlation between each feature and the outcome. Notably, *Mean rLSM* and *Semantic Relatedness* positively correlated with outcome, and *Max. Branching Factor Difference* correlated negatively. This is consistent with communication accommodation theory and shows that closer linguistic distances correlated with optimal outcomes.

The Glo. + Loc. model predicts interview success based on the interviewer's accommodation alone. To illustrate the affect of this accommodation on the success of an interview, in Figure 2 (c) we examine the relationship between predictions based on the Glo. + Loc. model and computational ground-truth when ranking the entire corpus by overall success. We can use this form of analysis to identify interviews where high levels of accommodation aligns with interview success (III), and where it does not (I). We can also spot interviews with a high success score despite a low level of measured accommodation (IX). We envision this inquiry helpful when a large corpus requires filtering before further analysis.

4. Conclusions

This paper introduces an automated cross-disciplinary approach for analysing interviews and successfully demonstrates it on a corpus of publicly available political interviews.

Our results confirm that we can successfully encode social scientific knowledge pertinent to interviewing into a computational analysis. Prudently, this can be harnessed both as a full analysis or as an initial mapping of a large corpus of conversational transcripts. Our method offers an interpretable and reproducible alternative to pre-labelled interview transcripts. This should encourage both computer scientists and social scientists alike when seeking to analyse conversations at scale.

Our psychologically-informed models significantly outperform a simple bag-of-words model, justifying domain-knowledge inclusion within computer science research. Using decipherable features also renders the analysis useful for future research within other domains. We have modelled an array of linguistic features, however, we note our choice of features is not exhaustive. Non-linguistic and paralinguistic behaviours may also contribute to accommodation.

Despite the close alignment between human and computer scores for specificity, our models did not successfully predict this measure based on the interviewer's behavior. This result may be specific to the political interview domain as establishing specific information is an unlikely goal within political interviewing. We choose to include specificity in this work as it may be of use for analysing interviews where the objective is more explicitly focused on information-gathering, for example, within the context of criminal investigation.

Our results indicate that political interviewing is a worthwhile setting to explore accommodation. A key advantage of our approach, however, is its transferability to other domains. We hope this work will lead to further adoption of the suggested cross-disciplinary approach to analyzing conversation at scale.

5. Acknowledgements

The authors would like to thank the participants who performed the manual validation task. We also thank the anonymous reviewers for providing valuable feedback. Darren Cook acknowledges support from the EPSRC and ESRC Centre for Doctoral Training in the Management and Quantification of Risk and Uncertainty in Complex Systems and Environments under ref. C110463K.

6. References

[1] D. Reitter and J. D. Moore, “Alignment and task success in spoken dialogue,” *Journal of Memory and Language*, vol. 76, pp. 29–46, 2014.

[2] I. Naim, M. I. Tanveer, D. Gildea, Mohammed, and Hoque, “Automated Analysis and Prediction of Job Interview Performance,”

IEEE Transactions of Affective Computing, vol. 9, no. 2, pp. 191–204, 2018.

[3] M. E. Ireland, R. B. Slatcher, P. W. Eastwick, L. E. Scissors, E. J. Finkel, and J. W. Pennebaker, "Language style matching predicts relationship initiation and stability," *Psychological Science*, vol. 22, no. 1, pp. 39–44, 2011.

[4] P. J. Taylor and S. Thomas, "Linguistic Style Matching and Negotiation Outcome," *Negotiation and Conflict Management Research*, vol. 1, no. 3, pp. 263–281, 2008.

[5] B. H. Richardson, P. J. Taylor, B. Snook, S. M. Conchie, and C. Bennell, "Language style matching and police interrogation outcomes," *Law and Human Behavior*, vol. 38, no. 4, pp. 357–366, 2014.

[6] H. H. Clark, *Using language*. Cambridge University Press, 1996.

[7] H. Giles, N. Coupland, and J. Coupland, "Accommodation theory: Communication, context, and consequence," in *Contexts of Accommodation*, H. Giles, N. Coupland, and J. Coupland, Eds. New York, NY: Cambridge University Press, 1991, ch. 1, pp. 1–68.

[8] A. Schweitzer, N. Lewandowski, and D. Duran, "Social attractiveness in dialogs," in *Proceedings of the 18th Annual Conference of the International Speech Communication Association*, F. Lacerda, D. House, M. Heldner, J. Gustafson, S. Strombergsson, and M. Wlodarczak, Eds. Stockholm, Sweden: International Speech Communication Association, 2017, pp. 2243–2247.

[9] A. Nenkova, A. Gravano, and J. Hirschberg, "High frequency word entrainment in spoken dialogue," in *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies Short Papers*, D. Lin, Y. Matsumoto, and R. Mihalcea, Eds. Columbus, OH: Association for Computational Linguistics, 2008, pp. 169–172.

[10] N. D. Duran, A. Paxton, and R. Fusaroli, "ALIGN: Analyzing Linguistic Interactions With Generalizable techNiques—A Python Library," *Psychological Methods*, vol. 24, no. 4, pp. 419–438, 2019.

[11] J. Rendle-Short, "Neutralism and adversarial challenges in the political news interview," *Discourse & Communication*, vol. 1, no. 4, pp. 387–406, 2007.

[12] J. Heritage, "Analyzing News Interviews: Aspects of the Production of Talk for an Overhearing Audience," in *Handbook of Discourse Analysis*, T. A. van Dijk, Ed. London, UK: Academic Press London, 1985, ch. 8, pp. 95–117.

[13] M. Waddle and P. Bull, "You're Important, Jeremy, but not that Important": Personalized Responses and Equivocation in Political Interviews," *Journal of Social and Political Psychology*, vol. 8, no. 2, pp. 560–581, 2020.

[14] B. Pluss, "Non-Cooperation in Dialogue," in *Proceedings of the ACL 2010 Student Research Workshop*, S. Demir, J. Raab, N. Reiter, M. Lopatkova, and T. Strzalkowski, Eds. Uppsala, Sweden: Association for Computational Linguistics, 2010, pp. 1–6. [Online]. Available: https://www.aclweb.org/anthology/P10-3000

[15] S. Young, "The broadcast political interview and strategies used by politicians: How the Australian prime minister promoted the Iraq War," *Media, Culture and Society*, vol. 30, no. 5, pp. 623–640, 2008.

[16] P. Grice, *Studies in the Way of Words*. Harvard University Press, 1989.

[17] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, and C. D. Manning, "Stanza : A Python natural language processing toolkit for many human languages," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, A. Celikyilmaz and T.-H. Wen, Eds. Seattle, WA: Association for Computational Linguistics, 2020, pp. 101–108.

[18] K. Collins and N. Carthy, "No rapport, no comment: The relationship between rapport and communication during investigative interviews with suspects," *Journal of Investigative Psychology and Offender Profiling*, vol. 16, no. 1, pp. 18–31, 2018.

[19] L. J. Alison, E. E. Alison, G. Noone, S. Elntib, and P. Christiansen, "Why tough tactics fail and rapport gets results: Observing rapport-based interpersonal techniques (ORBIT) to generate useful information from terrorists," *Psychology, Public Policy, and Law*, vol. 19, no. 4, pp. 411–431, 2013.

[20] M. Honnibal and I. Montani, "spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing," *To appear*, vol. 7, no. 1, pp. 411–420, 2017.

[21] M. Wilson, "MRC Psycholinguistic Database : Machine Usable Dictionary," *Behavior Research Methods, Instruments, & Computers*, vol. 20, no. 1, pp. 6–10, 1988.

[22] L. Zhou, J. K. Burgoon, J. F. Nunamaker, and D. Twitchell, "Automating Linguistics-Based Cues for detecting deception in text-based asynchronous computer-mediated communication," *Group Decision and Negotiation*, vol. 13, no. 1, pp. 81–106, 2004.

[23] R. Hou, V. Pérez-Rosas, S. Loeb, and R. Mihalcea, "Towards automatic detection of misinformation in online medical videos," in *2019 International conference on multimodal interaction*, 2019, pp. 235–243.

[24] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global Vectors for Word Representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, A. Moschitti, B. Pang, and W. Daelemans, Eds. Doha, Qatar: Association for Computational Linguistics, 2014.

[25] S. Garrod and M. J. Pickering, "Why is conversation so easy?" *Trends in Cognitive Sciences*, vol. 8, no. 1, pp. 8–11, 2004.

[26] W. Kulesza, D. Dolinski, A. Huisman, and R. Majewski, "The Echo Effect: The Power of Verbal Mimicry to Influence Prosocial Behavior," *Journal of Language and Social Psychology*, vol. 33, no. 2, pp. 183–201, 2014.

[27] T. L. Chartrand and J. A. Bargh, "The chameleon effect: the perception–behavior link and social interaction." *Journal of personality and social psychology*, vol. 76, no. 6, pp. 893–910, 1999.

[28] 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*, S. Sadagopan, K. Ramamritham, A. Kumar, and M. Ravindra, Eds. Hyderabad, India: Association for Computing Machinery, 2011, pp. 745–754.

[29] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, *The Development and Psychometric Properties of LIWC2015*. Austin, TX: Pennebaker Conglomerates, 2015.

[30] K. G. Niederhoffer and J. W. Pennebaker, "Linguistic style matching in social interaction," *Journal of Language and Social Psychology*, vol. 21, no. 4, pp. 337–360, 2002.

[31] L. C. Müller-Frommeyer, N. A. Frommeyer, and S. Kauffeld, "Introducing rLSM: An integrated metric assessing temporal reciprocity in language style matching," *Behavior Research Methods*, vol. 51, no. 3, pp. 1343–1359, 2019.

[32] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.