

Analyzing Discourse Structure and their Relationship to Dialogic Inertia in Facebook Pages of News Media Outlets

Mehtab Iqbal

Department of Information Technology

Georgia Southern University

Introduction

News media outlets provide provisions for users to interact with a news article through social media presence or directly, on their website (Hiller & Bakker, 2014; Diakopoulos & Naaman, 2011; Ruiz et al., 2011). Facebook is a popular social media where promoters and publishers can create pages to disseminate information to their users. Most news media outlets have a Facebook presence, where they share/post their articles on their individual page feed. Facebook users can then comment on these posts. These comments are considered to be important online participation by most news media outlets due to the wide reach of the platform (Oeldorf-Hirsch & Sundar, 2015).

On Facebook, these comments are structured at two levels. The original post can garner comments and, each individual comment can, in turn, be replied to by a user. This nested structure of comments provides a conducive environment for conversation, and debates (Iqbal & Khan, 2018). Several studies have found that up to 50% of comments have, at least, a single response (reply) made by another user (Ziegele, Breiner, & Quiring, 2014). According to Boczkowski & Mitchelstein (2012), these interactions are democratically valuable and contain rich interpersonal discourse on topics of public interest.

Discourse on Facebook pages of news media outlets tends to degenerate when two polarized communities interact with one another (Del Vicario, Zollo, Caldarelli, Scala, & Quattrociocchi, 2017). Users on Facebook tend to select the information that reinforces their confirmation bias and ignore dissenting information (Quattrociocchi, Scala, & Sunstein, 2016). This behavior catalyzes the formation of polarized groups (echo chamber) which further strengthens the community (Zollo et al., 2015b; Sunstein, 2002). We have seen in a previous study (Iqbal & Khan, 2018), interactions vary in different communities based on their partisan bias, and the partisan polarity of the community (news media followers).

These behavioral patterns can be found from within the linguistic components (Chung & Pennebaker, 2007) of each comment, and reply. Linguistic components, both syntactic and semantic, can elicit information on a person's mental state, emotions (Chung & Pennebaker, 2007), and social positioning (Milroy & Milroy, 1992). How people infer information, and then, in turn, respond to it takes place at a meta-cognitive level (Menyuk, 1985), which can be extracted from the metalinguistics characteristics of language (Gombert, 1992).

Metalinguistic analysis, not only helps us extract the surface realization of text but also provides us with the intentions of the interlocutor (writer) (Jakobson, 1963). Argumentation, a process of conveying opinion with the goal of making the opinion accepted (Peldszus & Stede, 2013), consists of various ways in which a premise changes in the course of the conversation to make complex information exchange (Freeman, 2011). To this end, Swanson, et al., (2015) have proposed an argument extraction algorithm based on the argumentative quality.

This argument quality relies on various linguistics, and metalinguistic hypotheses, and the richness of a sentence. Richness or the specificity is measured based on how specific, or generic a sentence is (Li and Nenkova, 2015). In this paper, we extracted the features based on the properties proposed by Swanson et al. (2015) and developed a framework for further analysis of what we label as the dialogic inertia within a discourse.

To form the basis of our dialogic inertia model, we posed three research questions based on how the specificity changes as argument progress. Finally, we would like to explore how a community's belief structure affects the direction, and inertia of a conversation.

RQ₁: *How does the richness (specificity) affect the continuance of the conversation?*

RQ₂: *How does the richness (specificity) change as conversation moves?*

RQ₃: *How does a community's political bias (alignment/misalignment) impact the length of the conversation and extent to which people share information?*

Literature Review

Metalinguistics

Coined around the 1950s, the term originally referred to 'metalanguage' - a language in its entirety consisted of linguistic terminology such as syntax, semantics, phoneme, lexeme, etc. In other words, metalanguage refers to a language whose sole purpose is to speak about language either formally (as in logic) and informally (Gombert, 1992). Language in itself is further broken down into two functions - principal, and the secondary (Jakobson, 1963). The principal functions concern itself with the three participants in a communication situation - the speaker, the addressee, and the subject which referred to in the message. The principal functions according to Jakobson (1963) can be broken down into three categories:

1. The expressive function (emotive function) determines the speaker's involvement in the communication event.
2. The connotative function is focused on how influential the message being transmitted to the partner.
3. The representative function (denotative function) which is centered around the content of the message.

The secondary functions are again categorized into three groups:

1. The phatic function revolves around the subtlety in communication which transcends the use of language and is used to keep the interlocutors in sync - for example, authoritarian discourse, polite exchange, etc.

2. The ludic and the poetic function is utilized when playing with language – for example, manipulation of words in poetry are carefully crafted grammatical alteration to imply emotions such as sarcasm, etc.
3. The metalinguistic function refers to the self-referential activity of dealing with the components of language and in turn, utilizes language as an object.

As Benveniste (1973) succinctly put it, metalinguistic ability refers to “the possibility of raising ourselves above language, of abstracting ourselves from it, of contemplating it, whilst making use of it in our reasonings and our observations” (Benveniste, 1973). This implies that dealing with metalinguistic activities is dependent on a cognitive effort that is beyond the strict boundaries of linguistic (Gombert, 1992).

Awareness of the structural characteristics of a language is a prerequisite for their intentional application by subjects in their own language-processing activities (‘we cannot use the knowledge which we do not have’) (Menyuk, 1985). Linguist identifies the ‘metalinguistic’ by examining verbal productions in order to find those linguistic features which indicate the existence of self-referential processes (the use of language to refer to itself) (Gombert, 1992). Psychologists, in contrast, analyze the behavior (verbal or otherwise) of the subject to discover elements which permit them to infer cognitive processes of conscious management (reflection on or intentional control over) of the language objects, either as objects or in terms of the use to which they are put (Gombert, 1992).

Emotional Content in written language

Empirical studies have shown that the way in which people use language can reveal information about their thoughts and emotions (Chung & Pennebaker, 2007). Linguistic Inquiry and Word Count (LIWC) has been successfully used to identify relationships between individuals in social interactions, including relative status (Sexton & Helmreich, 2000), deception (Newman, Pennebaker, Berry, & Richards, 2003), and the quality of close relationships (Slatcher & Pennebaker, 2006).

Social language processing is primarily concerned with inferring individual traits, such as sex, age, relative status, or mental health based on the use of language (Pennebaker, Mehl, & Niederhoffer, 2003). Sociolinguistics tends to combine social network analysis and linguistic style to determine social position using variation in linguistic patterns (Milroy & Milroy, 1992). For example, the use of pronouns provides information about how participants in a conversation address each other. First person singular pronoun usage tends to increase while someone interacts with a person of higher status (Kacewicz, Pennebaker, Davis, Jeon, & Graesser, 2014). The degree of emotionality can be determined from the use of social and affective words. Punctuations, referred to as discourse markers, can show how formal or informal the language being used is (Scholand et al., 2010).

Argument Extraction

In a social media context, arguments are unstructured and often, repetitive (Swanson, et al., 2015). Summarization of these arguments and the extraction of relevant argument statements is an important factor in further analyses of online argument and debate (Swanson, et al., 2015). To

this extent, Swanson, et al., (2015) posed a hypothesis termed as the “Implicit Markup Hypothesis”. The implicit markup hypothesis is a multifarious hypothesis which takes into consideration, several other hypotheses previously presented through various linguistic and metalinguistic understanding. The implicit markup hypothesis states that a good argument can be inferred as an argument from the surface realization of its linguistic components.

According to the “Discourse Relation Hypothesis”, arguments containing good argumentative segments tend to have an explicit connective between Arg1 and Arg2. Connectives are categorized into specification (“*first*”), contrast (“*but*”), contingency (“*if*”), and concession (“*so*”) (Prasad, et al., 2007). Additionally, the “Syntactic Properties Hypothesis” states that the syntactic properties of a sentence may provide a good argument indicator or even a sentential complement of a mental state (Marcu, 1999). The “Dialogue Structure Hypothesis” provides a strong relationship indicator between arguments by utilizing the dialogic structure of a conversation (Swanson, et al., 2015). In other words, collecting information on whether an argument is a direct reply to a statement or not is a straightforward feature of the extraction of argument statements. Finally, the “Semantic Density Hypothesis” states that the richness of a sentence is often a requirement towards the surface realization of an argument statement (Louis and Nenkova, 2011).

References

- Abbott, R., Walker, M., Anand, P., Fox Tree, J. E., Bowmani, R., & King, J. (2011). How Can You Say Such Things?!: Recognizing Disagreement in Informal Political Argument. In *Proceedings of the Workshop on Languages in Social Media* (pp. 2–11). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from <http://dl.acm.org/citation.cfm?id=2021109.2021111>
- Benveniste, E. (1973). Problems in General Linguistics (Miami Linguistic, No 8).
- Chung, C., & Pennebaker, J. W. (2007). The psychological functions of function words. *Social Communication, 1*, 343–359.
- Del Vicario, M., Zollo, F., Caldarelli, G., Scala, A., & Quattrociocchi, W. (2017). Mapping social dynamics on Facebook: The Brexit debate. *Social Networks, 50*, 6–16. <https://doi.org/10.1016/j.socnet.2017.02.002>
- Diakopoulos, N., & Naaman, M. (2011). Towards Quality Discourse in Online News Comments. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work* (pp. 133–142). New York, NY, USA: ACM. <https://doi.org/10.1145/1958824.1958844>
- Freeman, J. B. (2011). *Argument Structure:: Representation and Theory* (Vol. 18). Springer Science & Business Media.
- Galley, M., McKeown, K., Hirschberg, J., & Shriberg, E. (2004). Identifying Agreement and Disagreement in Conversational Speech: Use of Bayesian Networks to Model Pragmatic Dependencies. In *Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics. <https://doi.org/10.3115/1218955.1219040>
- Gombert, J. E. (1992). *Metalinguistic development*. University of Chicago Press.
- Hille, S., & Bakker, P. (2014). Engaging the Social News User: Comments on news sites and Facebook. *Journalism Practice, 8*(5), 563–572. <https://doi.org/10.1080/17512786.2014.899758>
- Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

- Iqbal, M., & Khan, S. (2018, March). Mining Facebook Page for Bi-Partisan Analysis. Paper presented at the SAIS, Atlanta, GA.
- Jakobson, R. (1963). Implications of language universals for linguistics. *Universals of Language*, 263–278.
- Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology*, 33(2), 125–143. <https://doi.org/10.1177/0261927X13502654>
- Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86.
- Li, J. J., & Nenkova, A. (2015). Fast and Accurate Prediction of Sentence Specificity. In *AAAI* (pp. 2281–2287).
- Louis, A., & Nenkova, A. (2011). Automatic identification of general and specific sentences by leveraging discourse annotations. In *Proceedings of 5th International Joint Conference on Natural Language Processing* (pp. 605–613). Chiang Mai, Thailand: Asian Federation of Natural Language Processing. Retrieved from <http://www.aclweb.org/anthology/I11-1068>
- Loper, E., & Bird, S. (2002). NLTK: The Natural Language Toolkit. In *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics - Volume 1* (pp. 63–70). Stroudsburg, PA, USA: Association for Computational Linguistics. <https://doi.org/10.3115/1118108.1118117>
- Marcu, D. (1999). Discourse trees are good indicators of importance in text. *Advances in Automatic Text Summarization*, 293, 123–136.
- Menyuk, P. (1985). Wherefore Metalinguistic Skills? A Commentary on Bialystok and Ryan. *Merrill-Palmer Quarterly*, 31(3), 253–259.
- Milroy, L., & Milroy, J. (1992). Social network and social class: Toward an integrated sociolinguistic model. *Language in Society*, 21(1), 1–26.
- Misra, A., & Walker, M. (2017). Topic Independent Identification of Agreement and Disagreement in Social Media Dialogue. *ArXiv:1709.00661 [Cs]*. Retrieved from <http://arxiv.org/abs/1709.00661>
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29(5), 665–675.

- Oeldorf-Hirsch, A., & Sundar, S. S. (2015). Posting, commenting, and tagging: Effects of sharing news stories on Facebook. *Computers in Human Behavior*, 44, 240–249. <https://doi.org/10.1016/j.chb.2014.11.024>
- Otero, V. (2018, February 5). The Chart, 3.1 Minor Updates Based on Constructive Feedback. Retrieved May 4, 2018, from <http://www.allgeneralizationsarefalse.com/chart-3-1-minor-updates-based-constructive-feedback/>
- Peldszus, A., & Stede, M. (2013). From argument diagrams to argumentation mining in texts: A survey. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, 7(1), 1–31.
- Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54(1), 547–577.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9), 3526–3529. <https://doi.org/10.1073/pnas.1012551108>
- Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., Joshi, A., Robaldo, L., & Webber, B. (2007). The Penn Discourse Treebank 2.0 Annotation Manual. IRCS Technical Reports Series. Retrieved from https://repository.upenn.edu/ircs_reports/203
- Quattrociocchi, W., Scala, A., & Sunstein, C. R. (2016). Echo chambers on facebook
- Ruiz, C., Domingo, D., Micó, J. L., Díaz-Noci, J., Meso, K., & Masip, P. (2011). Public Sphere 2.0? The Democratic Qualities of Citizen Debates in Online Newspapers. *The International Journal of Press/Politics*, 16(4), 463–487. <https://doi.org/10.1177/1940161211415849>
- Scholand, A. J., Tausczik, Y. R., & Pennebaker, J. W. (2010). Social language network analysis. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (pp. 23–26). ACM.
- Sexton, J. B., & Helmreich, R. L. (2000). Analyzing cockpit communications: the links between language, performance, error, and workload. *Human Performance in Extreme Environments*, 5(1), 63–68.
- Sunstein, C. R. (2002). The law of group polarization. *Journal of Political Philosophy*, 10(2), 175–195.
- Slatcher, R. B., & Pennebaker, J. W. (2006). How do I love thee? Let me count the words: The social effects of expressive writing. *Psychological Science*, 17(8), 660–664.

- Swanson, R., Ecker, B., & Walker, M. (2015). Argument mining: Extracting arguments from online dialogue. In Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue (pp. 217–226).
- Walker, M. A., Tree, J. E. F., Anand, P., Abbott, R., & King, J. (2012). A Corpus for Research on Deliberation and Debate. In *LREC* (pp. 812–817).
- Wang, L., & Cardie, C. (2016). Improving Agreement and Disagreement Identification in Online Discussions with A Socially-Tuned Sentiment Lexicon. *ArXiv:1606.05706 [Cs]*. Retrieved from <http://arxiv.org/abs/1606.05706>
- Ziegele, M., Breiner, T., & Quiring, O. (2014). What Creates Interactivity in Online News Discussions? An Exploratory Analysis of Discussion Factors in User Comments on News Items. *Journal of Communication*, 64(6), 1111–1138. <https://doi.org/10.1111/jcom.12123>
- Zollo, F., Novak, P. K., Del Vicario, M., Bessi, A., Mozetič, I., Scala, A., ... Quattrociocchi, W. (2015). Emotional Dynamics in the Age of Misinformation. *PLOS ONE*, 10(9), e0138740. <https://doi.org/10.1371/journal.pone.0138740>