

Capturing Turn-by-Turn Lexical Similarity in Text-Based Communication

Noah Liebman
Northwestern University
Evanston, IL, USA
nliebman@u.northwestern.edu

Darren Gergle
Northwestern University
Evanston, IL, USA
dgergle@northwestern.edu

ABSTRACT

Speakers often come to use similar words during conversation; that is, they come to exhibit *lexical similarity*. The extent to which this occurs is associated with many positive social outcomes. However, existing measures of lexical similarity are either highly labor intensive or too coarse in their temporal resolution. This limits the ability of researchers to study lexical similarity as it unfolds over the course of a conversation. We present a fully automated metric for tracking lexical similarity over time, and demonstrate it on individual conversations, explore general trends in aggregate conversational dynamics, and examine differences in how similarity tracks over time in groups with differing social outcomes.

Author Keywords

Conversation; similarity; lexical entrainment; collaboration; coordination; methodology

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

Lexical entrainment [10], or the process by which people converge upon the terms they use to establish shared conceptualizations and perspectives, is a well-established conversational phenomenon that has a wide range of collaborative benefits. Its presence is associated with a number of positive social outcomes ranging from how much people like one another [4] to increased feelings of attractiveness [17] to better coordination on tightly-coupled group tasks [11]. These benefits are not limited to face-to-face interactions: lexical similarity has been shown to have similar advantages in technologically mediated settings. It is associated with task success [18], the self-reported quality of an interaction [19], trust [22], and group cohesiveness [12].

However, while lexical similarity is a well-established and socially advantageous phenomenon, we currently lack efficient

and effective ways to track its progress during interactions that take place in online communities, instant messaging, or messaging apps like SMS or WhatsApp. Many techniques rely on time-consuming hand-coding, and current automated methods provide coarse results at the level of entire conversations. As a result we miss out on analytical opportunities as well as opportunities for novel technology design.

We aim to address these shortcomings by developing a technique that allows researchers to *automatically capture lexical similarity as it unfolds over a conversation*. To illustrate our approach, consider the following scenario (illustrated in Figure 1) that demonstrates one of the many uses of our technique:

Two friends, Alice and Bob, are engaged in an instant message conversation. It is important to keep in mind that at any time they can express a thought to one another in a nearly infinite number of ways. Yet, as they converse, they influence each other's word choice [3], and as a result we see greater amounts of lexical similarity. For example, Alice and Bob greet each other using the exact same word. They both use "Hey" instead of one person saying "Hey" and the other saying "Hi", which results in a peak in our measure near the beginning of the graph. As they begin to talk about Alice's broken necklace, we see that clarifying where the necklace broke results in an uptick in similarity. This stems from repeated use of the term "the chain" even when a pronoun would suffice. We then come to an example of repeating a partner's words even when not necessary, presumably because it serves a social function. In the discussion about vices, Bob could have just asked, "What's that?", but he instead repeated "Veronica Mars" along with other of Alice's words. Finally, we see lexical similarity resulting from similar valedictions, both containing "night".

This example scenario shows how our measure responds to local changes in lexical similarity — detecting greater similarity when interlocutors have more words in common within a local window — and it illustrates how this can be used to maintain a running record and highlight areas of conversation where lexical similarity occurs.

In this paper, we introduce our measure, detail its calculation, and demonstrate its utility across three potential applications: measurement of individual conversations, measurement of aggregated conversations, and its potential use for detecting statistical differences across sets of conversations.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
CSCW '16, February 27–March 2, 2016, San Francisco, CA, USA
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3592-8/16/02...\$15.00
DOI: <http://dx.doi.org/10.1145/2818048.2820062>

Through these examples we hope to give researchers an idea of the applicability of our measure to the computational social sciences.

BACKGROUND

When people communicate successfully, their language and style tend to be similar [5]. In this work we focus on lexical similarity¹, or the extent to which the speakers use the same words and phrases. For example, if one person refers to the final meal of the day as “dinner” and the other calls it “supper”, their language is lexically dissimilar, whereas if they both use the term “supper”, their language is lexically similar.

This sort of similarity can be explained by a number of theories, including Communication Accommodation Theory (CAT) [9], Grounding [6], and Linguistic Style Matching [19]. While these theories assume different social or psychological mechanisms (see [24] for a review), they can result in similar surface phenomena. Our goal with this work is to create a tool social scientists can use to capture and study such phenomena in order to advance the state of social science theory.

Measures of similarity can be an important tool for researchers studying collaboration and coordination, and automating such measures offers a number of benefits, including scalability, which enables efficient large-scale corpus analyses. With so much communication taking place in computer-mediated environments, this is a potentially powerful capability, as we will demonstrate. Additionally, lexical similarity is recognized as an important component for automated dialog systems and voice user interfaces [2], so the ability to assess similarity in real time can have benefits for such systems.

Most existing automated methods for assessing similarity aim to derive a single measure for an entire conversation; that is, they consider global, but not local, similarity. In addition, they often use a combination of semantic and lexical analyses (e.g., [19, 12]). Such work often involves a degree of hand coding, and combines this with automated semantic tools such as LIWC [20] (e.g., [21]). For their turn-by-turn and conversation-level analyses, [19] used a comparison of words based on LIWC categories (e.g., emotion words, past tense verbs) rather than matching on individual words and word forms. The Linguistic Style Matching (LSM) metric proposed by [12] is similar, relying on LIWC categories and function words. Scissors et al. [22] took a hand-coding approach to measuring similarity along lexical, syntactic, and semantic dimensions. They later augmented this with automated analyses using LIWC categories and an analysis that combined identifying phrases with an automatic parser with hand coding of those phrases [21].

Lexical entrainment can operate either *historically* or *ahistorically*, depending on whether convergence on a given word is because that word was used previously in the conversation

¹We focus on lexical similarity as opposed to semantic similarity, which typically captures whether interlocutors are discussing the same topic, or syntactic similarity, which considers whether interlocutors use similar sentence structures.

or whether it is simply the most probable word choice [3]. While studies have found general support for historical models of lexical entrainment (e.g., [10, 25, 26]), there are no existing metrics to assess lexical similarity over time in a fully automatic way.

One measure that does consider local variation in similarity over time is Discursis [1]. Discursis provides a measure and visualization of topical coherence over time. The similarity measure is semantic, although its semantic model is based on lexical term co-occurrence. The visualization uses a grid of colored squares to draw attention to conceptual similarity between pairs of utterances, even if they are not adjacent. This differs from our measure, which yields a time-varying measure of similarity rather than a map of topical coherence.

Closer to the work presented here, recent work by Foltz and colleagues has shown promise using more purely lexical analysis [8], especially as there is evidence that syntactic similarity is highly associated with lexical similarity [14]. They used cosine similarity, a mathematical method to calculate lexical similarity between two speakers. Their word frequency vectors only included words that were automatically tagged as nouns, adjectives, or verbs. In addition, they also took some first steps to consider how similarity varies over the course of an interaction. They did this by calculating similarity after each full round of conversation in an experiment with multiple rounds, and showed how this could provide insight into lexical entrainment. By considering overall similarity after each round of a multiple-round interaction, Foltz’s [8] automatic analysis comes closer to capturing lexical entrainment over time, but still considers larger blocks of time. Our technique builds on this approach and provides a flexible and fine-grained method for examining lexical entrainment as it unfolds over a conversation on a turn-by-turn basis.

OUR APPROACH

We present a metric for tracking lexical similarity as it varies over the course of a conversation. This metric relies strictly on lexical features, independent of part of speech, semantic meaning, or other non-lexical consideration. Additionally, it accounts for baseline levels of lexical similarity that arise from the joint use of frequent words and other ahistorical sources. In assessing our approach, we find evidence that lexical similarity has temporal structure significantly greater than chance even after accounting for high-frequency tokens.

METHOD

Influenced by [8], the similarity metric presented here is an application of cosine similarity to language analysis [7]. Cosine similarity is a general measure of the similarity between two vectors, \mathbf{A} and \mathbf{B} , of arbitrary but equal dimensionality. Similarity is the cosine of the angle between the two vectors, θ , which, by the definition of the dot product, is equal to Equation 1.

Data Pre-Processing

To evaluate our metric, we made use of chat (instant messaging) data. To transform the chat data into multidimensional vectors, we need both dimensions (i.e., directions) and magnitudes. Dimensions represent tokens, which in this case are

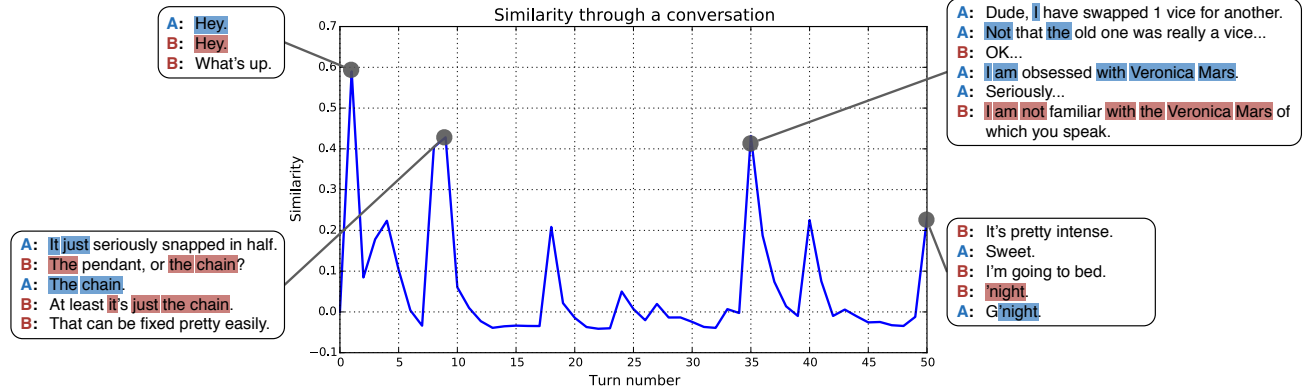


Figure 1. Example conversation between Alice and Bob

$$\begin{aligned}
 \text{similarity}(\mathbf{A}, \mathbf{B}) &= \cos \theta \\
 &= \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} \\
 &= \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}}.
 \end{aligned} \tag{1}$$

words, contractions, and other punctuation of interest. We made sure that each token was all lowercase, then removed all spaces, commas, periods, and question marks, but retained all other punctuation, including exclamation points and emoticons. Unlike [8], we did not limit our analysis to certain parts of speech because both content words (e.g., [8]) and function words (e.g., [12]) are known to be important. Each magnitude is a weighted term frequency, the calculation of which we will now describe.

Calculating Similarity

We computed a vector of weighted term frequencies at each turn in a conversation. A turn is a transmission or series of consecutive transmissions by one sender. In instant message conversations, it is common for one person to hold the floor for multiple consecutive transmissions [16]. Therefore, we treat a full turn as a single point in time, even if it spans multiple transmissions.

Windowing

Conversations unfold over time, so we used a sliding window to evaluate local similarity. We used a linear, four-turn-wide window that trailed turn t (Equation 2), the value of which was used to weight the relevant term frequencies.

$$\text{window}(t) = -\frac{1}{4}t + 1 \tag{2}$$

This decay function had the effect of building in a recency bias, which is our goal if we want to measure temporally local similarity. Words used at time t are given full weight, while

those at $t - 1$ to $t - 3$ are given progressively less weight. An extended and parameterizable window is important because there is evidence that typical forms of topical coherence in spoken conversation are regularly violated in computer-mediated communication [13].

Inverse document frequency

One way we account for similarity that arises from both interlocutors speaking the same language (i.e., ahistorical similarity) is by accounting for common, or high frequency, words such as “a” or “the”. To do this, we used term frequency-inverse document frequency (TF-IDF) [23]. TF-IDF is typically used in applications with sets of large documents to determine whether a term is important to a particular document. IDF weights term frequencies within a document by a value that is inversely proportional to the total number of documents that term appears in.

In our case, the raw values in \mathbf{A} and \mathbf{B} are already term frequencies (TF), but we need to define the IDF component. We define a document as a single turn t within a full conversation, T , of N turns. Each turn contains words which may or may not be present in all turns. We therefore define IDF^2 :

$$\begin{aligned}
 \text{IDF}(w, T) &= \log \frac{N}{1 + |\{t \in T : w \in t\}|} \\
 &= \log \frac{\text{number of turns in } T}{1 + \text{number of turns containing term } w}.
 \end{aligned} \tag{3}$$

The IDF weight of a word can then be used to weight that word’s frequencies in the similarity calculation (Equation 4).

We do not wish to entirely discount high-frequency words, however, because similarity in function words, many of which occur frequently, can be predictive of social outcomes [12].

²It is common to add 1 to the denominator to avoid dividing by 0 in cases when a term is not present in any documents.

$$\text{similarity}(\mathbf{A}, \mathbf{B}, t) = \frac{\sum_{i=t-3}^t (\text{window}(t-i))^2 A_i \text{IDF}(A_i, \mathbf{A} \cup \mathbf{B}) B_i \text{IDF}(B_i, \mathbf{A} \cup \mathbf{B})}{\sqrt{\sum_{i=t-3}^t [\text{window}(t-i) A_i \text{IDF}(A_i, \mathbf{A} \cup \mathbf{B})]^2} \times \sqrt{\sum_{i=t-3}^t [\text{window}(t-i) B_i \text{IDF}(B_i, \mathbf{A} \cup \mathbf{B})]^2}} \quad (4)$$

Permutation analysis

We are interested specifically in similarity that develops over time. To account for other forms of time-independent similarity, either from the language itself or from artifacts of the analysis, we performed a permutation analysis and computed a baseline atemporal similarity for each time point, then subtracted it out of the true similarities.

We shuffled the order of the turns in a conversation 1,000 times³, ran the shuffled conversations through the analyses described above (Equation 4), and took the mean of those 1,000 iterations at each turn. After subtracting out this baseline level of similarity, we were left with just similarity that arises from temporal similarity between interlocutors.

RESULTS

In the following sections we provide examples of the sorts of analysis our method enables. While they may not have the depth of full analyses, the intent is for them to be illustrative of the utility of the technique we have just described.

Data

The data used here comes from a corpus of 60 instant message conversations collected in a laboratory setting. Pairs did not know each other, and were given the task of trying to come to an agreement on an assigned dilemma. Half of the pairs were given 15 minutes to chat, while the other half were given five minutes, resulting in 23,467 words across 2,363 lines chat text. Participants had a mean age of 22.9 years, and there were 30 female–female, 25 male–female, and five male–male pairs. For more details about the data collection process, see [15].

Individual Conversations

One way of applying this measure is to gain insight into phenomena occurring in individual conversations. We envision this as part of a mixed-methods, qualitative/quantitative analysis of conversational data. In this example, we can see how the structure of conversations comes through in both the plotted similarity and the simple qualitative descriptions.

The somewhat out-sized influence of greetings on similarity toward the beginning is evident in all of the example conversations shown here (Figure 2). This is because the total number of turns (and words) is limited by the window being wider than the number of turns that have occurred so far.

In the conversation in Figure 2A, after their greeting, the pair settles on some common vocabulary for discussing the situation around turn five (e.g., “think”, “tell”), then express disagreement with each other for much of the middle part of the

³We used visual inspection to determine how many iterations were required for the baseline to converge.

conversation (turns 15–25 or so). Finally, the pair begins to come to an agreement around turn 28 before ending somewhat abruptly with a quick exchange of “bye”s.

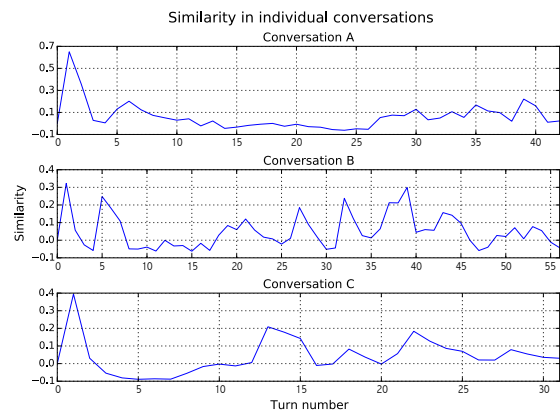


Figure 2. Similarity for three different individual conversations

The conversation in Figure 2B contains a lot of arguing back and forth, and features a number of phrases that are repeated nearly verbatim by both partners. For example, “well his parents *will be mad at him* of course” is answered with “They *will be mad at him* when they find out”; “Why not *lie for another month or two* and then confess?” is answered with “Do you REALLY think *lying for another month or two* will make them happy?”. The conversation ends with no sign-off when time expired.

Finally, the conversation in Figure 2C starts with a typical mutually common greeting (“hey”). They then take a bit of time to get situated in the scenario they are supposed to be discussing, then start using more topical words like “sibling” around turn 11. As with the previous conversation, they end with no sign off.

Aggregated Conversations

We also applied this metric to all 60 conversations in our corpus at once. The following figures show the mean similarity aggregated across all conversations over time; that is, the similarity for each conversation scaled to a uniform length.

When scaled to a uniform length, the notions of “turn” and “time” are made more complicated. When considering individual conversations, each conversational turn represented a single point in time; the two concepts are indistinguishable in our measure. However, when conversations have different numbers of turns, the two concepts become uncoupled.

As in the previous section, turns represent actual conversational turns in the data, but here time is a proportion of the

way through a conversation, independent of the number of turns. This is so conversations with different numbers of turns can be compared and aggregated more directly. Before aggregation, each conversation’s results were piecewise constant interpolated to a common length (effectively 100 turns).

The following figures show the mean similarity across all conversations over time; that is, the similarity for each conversation is scaled to a uniform length, and the mean is taken across all 60 conversations.

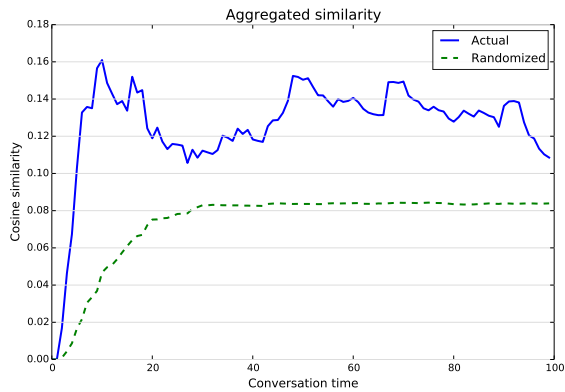


Figure 3. Cosine similarity over time and its randomized baseline

To point out the influence of the permutation analysis, Figure 3, “Randomized”, shows the baseline itself, while Figure 3, “Actual”, shows the result of Equation 4 only. We see that the baseline from the permutation analysis captures the ramp-up due to the window at the beginning of conversations, but quickly converges to a steady state. The steady level of similarity represents a background level of similarity between partners independent of temporal effects. The actual similarities are always greater than the baseline. This indicates that similarity is at least partially driven by temporal processes.

The final similarity measure is the difference between the similarity calculated in Equation 4 and the randomized baseline. Figure 4 shows that final measure over time after accounting for similarity from high frequency words and non-temporal structure. When looking at the 95% confidence interval, we see that it is always greater than zero. This indicates that entrainment through time plays a significant role in establishing lexical similarity; that is, we find evidence that lexical entrainment operates historically.

This also reveals that the significant peak toward the beginning of conversations is not just an artifact of the analysis. Although partially because there are fewer words in the window at the beginning of the conversation, we also believe this structure is a result of using common greeting words such as “hi” and “hey” in the first few turns.

Experimental Application

Finally, to demonstrate how this metric may be used in an experimental setting, we assessed whether there were significant differences in the similarities between pairs who rated their interactions as having a high degree of affinity and those

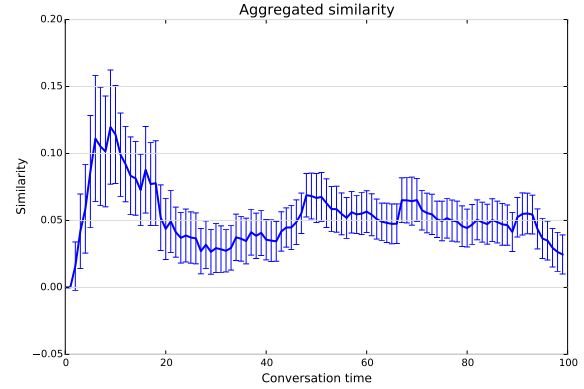


Figure 4. Final similarity metric (above randomized baseline) aggregated across conversations

who rated them as having low affinity (Figure 5). In this example, pairs are divided into the top and bottom thirds of affinity ratings. (For more details on these ratings, see [15].)

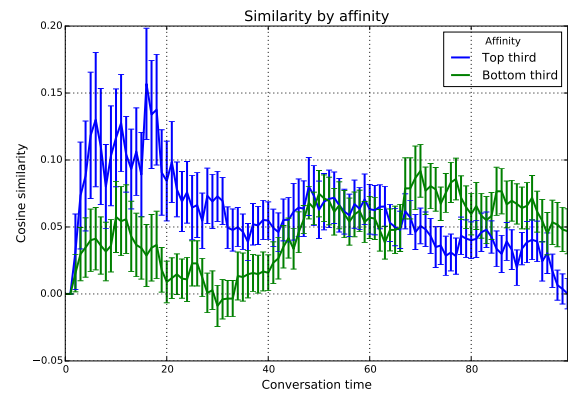


Figure 5. Similarity aggregated across conversations in the top and bottom tertile of affinity rating

The unique advantage of our metric is that it can reveal trends that unfold over the course of a conversation. Here, our results suggests that high-affinity pairs tend to start conversations with significantly greater similarity, while low-affinity pairs finish their conversations with greater similarity. Exploring the mechanisms behind these dynamics will require future work, as our aim in this paper to introduce this similarity metric.

DISCUSSION

We have introduced a method for measuring lexical similarity over time. It uses cosine similarity, with weighting to account for high-frequency words, and compensation for non-temporal sources of lexical similarity. A sliding window reveals similarity that arises from local conversational structure. When aggregated across many conversations, this can only show more macro-scale similarity tendencies, but for a single conversation it may reveal smaller-scale social and linguistic processes throughout a conversation.

The measure itself is quite flexible, both in how it works and potential applications. Window width and shape are easily parametrizable. The window we used was small and linear, but one could imagine wider windows with different envelopes. As the individual example conversations showed, the four-turn-wide window allows us to capture highly localized similarity, but may miss broader trends. Another possibility is to dynamically size a window based on a better understanding of topical coherence.

How we account for high frequency words is also flexible. We defined the set of all documents for the purposes of calculating IDF weights by essentially asking, “How important is this word to the conversation?” It would also be interesting to see how defining the set of documents in terms of how important a word is to the speaker or listener of a given turn affects the computed similarity. This could help give insight into differences in how individuals adapt to the language of another, and such analyses could be particularly useful for examining patterns of influence or dominance in negotiation settings. Alternatively, assessing word frequency based on a larger, more global corpus would likely provide a more accurate picture of which words are truly high frequency in a language, in addition to the advantage of enabling real-time measurement.

Depending on how these methods are applied, it may be desirable to track lexical similarity in real time, as a conversation is taking place. As mentioned above, calculating term frequencies using a global, rather than conversation-specific, corpus would enable real time term frequency weighting. It is not possible to run the permutation analysis without knowing the entire conversation up front, though as we saw in Figure 3, it is essentially a constant that gets subtracted, the value of which it may be possible to estimate.

The analyses shown here have been primarily quantitative, but pairing them with a qualitative analysis of conversations would add substantial depth. We hinted at this in an informal way for a small number of conversations, but we believe our measure could be a valuable part of a more rigorous mixed methods approach to understanding social dynamics in conversations.

For researchers considering using an automated measure of similarity, understanding the differences between the measure we presented in this paper and existing measures is important. The attribute of our measure that makes it most distinct is the fact that it results in time-series, turn-by-turn data. By contrast, LSM [12], the measure developed by Foltz [8], and others result in a single value for an entire conversation. Discursis [1] takes a third approach, giving insight into time-based phenomena not in a strictly linear sense, as we do, but pairwise by turns. If the question under study is about how similarity changes or develops over time, our measure is more appropriate; questions of topical coherence and structure may be better addressed with a method like Discursis.

Another unique element of our measure is that it is fully lexical. LSM and Discursis both use semantic models, which adds a level of complexity that, depending on the theoretical

orientation of a researcher, may not be desired. Foltz et al.’s, like our measure, is also lexical, but they only considered certain words (nouns, adjectives, and verbs) rather than all utterances. Using all utterances can be particularly advantageous in a computer-mediated setting because many text tokens are not words (e.g., punctuation and emoticons).

While Foltz only considered certain words and made manual corrections to spelling, we took a more naïve approach in order to demonstrate our method’s potential applicability to big data applications; however, these decisions are more a question of data pre-processing than computational method. In terms of computational method, we did introduce a sliding window, automatic weighting of highly frequent terms, and a permutation analysis to isolate temporal effects.

Finally, although we did not set out to do so, we also make a modest contribution to theory. In the aggregate, temporal lexical similarity is consistently significantly greater than zero, even after subtracting out atemporal sources of similarity. This provides evidence for historical lexical entrainment, as similar words tend to be used within relatively close proximity of each other.

CONCLUSION

With this work we make two primary contributions. First, we introduce a metric that tracks the lexical similarity between two interlocutors and how it evolves over the course of a conversation. Second, we demonstrate three potential applications of such a measure: as an aid in mixed-methods analysis of individual conversations, to aggregate conversations to assess general trends, and in an experimental setting to reveal differences between groups. We believe that our metric is flexible and will be a useful tool for researchers in the CSCW community.

ACKNOWLEDGEMENTS

We thank the reviewers for their insightful feedback on this work. This work was funded by National Science Foundation grant #0953943.

REFERENCES

1. Daniel Angus, Andrew Smith, and Janet Wiles. 2012. Conceptual Recurrence Plots: Revealing Patterns in Human Discourse. *IEEE Transactions on Visualization and Computer Graphics* 18, 6 (June 2012), 988–997. DOI : <http://dx.doi.org/10.1109/TVCG.2011.100>
2. Susan E. Brennan. 1998. The Vocabulary Problem in Spoken Dialogue Systems. In *Automated Spoken Dialog Systems*, Luperfoy (Ed.). MIT Press, Cambridge, MA.
3. Susan E. Brennan and Herbert H. Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22, 6 (1996), 1482–1493. <http://dx.doi.org/10.1037/0278-7393.22.6.1482>
4. Tanya L. Chartrand and John A. Bargh. 1999. The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology* 76, 6 (1999), 893–910. DOI : <http://dx.doi.org/10.1037/0022-3514.76.6.893>

5. Herbert H. Clark. 1996. *Using Language*. Cambridge University Press, Cambridge.
6. Herbert H. Clark and Susan E. Brennan. 1991. Grounding in Communication. In *Perspectives on Socially Shared Cognition*, Lauren B Resnick, John M Levine, and Stephanie D Teasley (Eds.). American Psychological Association, 127–149.
7. Marc Damashek. 1995. Gauging Similarity with n-Grams: Language-Independent Categorization of Text. *Science* 267, 5199 (Feb. 1995), 843–848. DOI : <http://dx.doi.org/10.1126/science.267.5199.843>
8. Anouschka Foltz, Judith Gaspers, Carolin Meyer, Kristina Thiele, Philipp Cimiano, and Prisca Stenneken. 2014. Temporal Effects of Alignment in Text-Based, Task-Oriented Discourse. *Discourse Processes* (Dec. 2014). DOI : <http://dx.doi.org/10.1080/0163853X.2014.977696>
9. Cindy Gallois, Tania Ogay, and Howard Giles. 2005. Communication accommodation theory: A look back and a look ahead. In *Theorizing about intercultural communication*, W B Gudykunst (Ed.). Sage, Thousand Oaks, CA, 121–148.
10. Simon Garrod and Anthony Anderson. 1987. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition* 27, 2 (Nov. 1987), 181–218. DOI : [http://dx.doi.org/10.1016/0010-0277\(87\)90018-7](http://dx.doi.org/10.1016/0010-0277(87)90018-7)
11. Darren Gergle, Robert E. Kraut, and Susan R. Fussell. 2013. Using Visual Information for Grounding and Awareness in Collaborative Tasks. *Human-Computer Interaction* 28, 1 (2013), 1–39. DOI : <http://dx.doi.org/10.1080/07370024.2012.678246>
12. Amy L. Gonzales, Jeffrey T. Hancock, and James W. Pennebaker. 2010. Language Style Matching as a Predictor of Social Dynamics in Small Groups. *Communication Research* 37, 1 (Jan. 2010), 3–19. DOI : <http://dx.doi.org/10.1177/0093650209351468>
13. Susan Herring. 1999. Interactional Coherence in CMC. *Journal of Computer-Mediated Communication* 4, 4 (1999). DOI : <http://dx.doi.org/10.1111/j.1083-6101.1999.tb00106.x>
14. Christine Howes, Patrick G.T. Healey, and Matthew Purver. 2010. Tracking Lexical and Syntactic Alignment in Conversation. *Proceedings of the Annual Meeting of the Cognitive Science Society* (2010).
15. Noah Liebman and Darren Gergle. 2016. It's (Not) Simply a Matter of Time: The Relationship Between CMC Cues and Interpersonal Affinity. *CSCW '16: Proceedings of the 2016 ACM Conference on Computer Supported Cooperative Work* (2016). DOI : <http://dx.doi.org/10.1145/2818048.2819945>
16. Rich Ling and Naomi S. Baron. 2007. Text Messaging and IM: Linguistic Comparison of American College Data. *Journal of Language and Social Psychology* 26, 3 (2007), 291–298. DOI : <http://dx.doi.org/10.1177/0261927X06303480>
17. Clifford Nass and Kwan Min Lee. 2000. Does computer-generated speech manifest personality? an experimental test of similarity-attraction. In *CHI '00*. ACM Press, New York, New York, USA, 329–336. DOI : <http://dx.doi.org/10.1145/332040.332452>
18. Ani Nenkova, Agustín Gravano, and Julia Hirschberg. 2008. High frequency word entrainment in spoken dialogue. *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers* (June 2008), 169–172. <http://dl.acm.org/citation.cfm?id=1557690.1557737>
19. Kate G. Niederhoffer and James W. Pennebaker. 2002. Linguistic Style Matching in Social Interaction. *Journal of Language and Social Interaction* 21 (2002), 337–360. <http://jls.sagepub.com/content/21/4/337.short>
20. James W. Pennebaker, Roger J. Booth, and Martha E. Francis. 1999. Linguistic Inquiry and Word Count (LIWC). (1999). <http://www.liwc.net/>
21. Lauren E. Scissors, Alastair J. Gill, Kathleen Geraghty, and Darren Gergle. 2009. In CMC we trust: the role of similarity. *CHI '09: Proceedings of the 27th International Conference on Human Factors in Computing Systems* (April 2009). DOI : <http://dx.doi.org/10.1145/1518701.1518783>
22. Lauren E. Scissors, Alastair J. Gill, and Darren Gergle. 2008. Linguistic mimicry and trust in text-based CMC. *CSCW '08: Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (Nov. 2008). DOI : <http://dx.doi.org/10.1145/1460563.1460608>
23. Karen Spärck Jones. 1972. A Statistical Interpretation of Term Specificity and its Application in Retrieval. *Journal of Documentation* 28, 1 (Jan. 1972), 11–21. DOI : <http://dx.doi.org/10.1108/eb026526>
24. Catalina L. Toma. 2014. Towards Conceptual Convergence: An Examination of Interpersonal Adaptation. *Communication Quarterly* 62, 2 (April 2014), 155–178. DOI : <http://dx.doi.org/10.1080/01463373.2014.890116>
25. Mija M. Van Der Wege. 2009. Lexical entrainment and lexical differentiation in reference phrase choice. *Journal of Memory and Language* 60, 4 (May 2009), 448–463. DOI : <http://dx.doi.org/10.1016/j.jml.2008.12.003>
26. Si On Yoon and Sarah Brown-Schmidt. 2013. Lexical differentiation in language production and comprehension. *Journal of Memory and Language* 69, 3 (Oct. 2013), 397–416. DOI : <http://dx.doi.org/10.1016/j.jml.2013.05.005>