

A Dynamical Approach to Extracting Language Style Patterns using Recurrence Quantification

Aaron Necaie

anecaie@iu.edu

Indiana University Bloomington <https://orcid.org/0000-0002-1111-2898>

Jingjing Han

Fudan University

Mary Jean Amon

Indiana University Bloomington

Research Article

Keywords: Language style matching, language synchrony, recurrence quantification analysis, dynamical systems, communication dynamics

Posted Date: September 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11118758/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

A Dynamical Approach to Extracting Language Style Patterns using Recurrence Quantification

Aaron Necaie^{1*}, Jingjing Han² and Mary Jean Amon³

¹Luddy School of Informatics, Computing, and Engineering, Indiana University, 700 N. Woodlawn Avenue, Bloomington, 47408, USA.

²School of Journalism, Institute for Global Communications and Integrated Media, Fudan University, 440 Handan Rd, Shanghai, 200437, China.

*Corresponding author(s). E-mail(s): anecaais@iu.edu;
Contributing authors: [Aaron Necaie Scholar Profile](#); [Jingjing Han Scholar Profile](#); [Mary Jean Amon Scholar Profile](#);

Abstract

Language style matching (LSM) is a type of linguistic synchrony in which conversation partners mimic each other’s grammatical structure, as opposed to semantic meaning, where LSM is linked to a variety of prosocial outcomes like empathy and positive affect. However, the traditional LSM approach provides global estimates of synchrony aggregated across the conversation, reducing fidelity and limiting understandings of the time-evolving patterns that define language. For example, there are challenges in examining synchrony at the turn-level, because of the difficulty in accounting for reciprocal and nonlinear conversation structures. The present work introduces a comprehensive framework for dynamically studying turn-level LSM through integration of recurrence quantification analysis (RQA), creating a hybridized approach referred to as recurrence-based LSM (rbLSM). This technique identifies temporal linguistic patterns between dyads, multi-speaker groups, or within single-speaker monologues. A series of examples demonstrate how rbLSM quantifies dynamic patterns at turn, speaker, and conversation levels. Applying rbLSM to facial electromyography (EMG) recordings from dyads, we further show how rbLSM metrics predict valued outcomes, in this case, positive emotional expressions during conversation. Limitations and practical considerations are discussed.

Keywords: Language style matching, language synchrony, recurrence quantification analysis, dynamical systems, communication dynamics

1 Introduction

It has long been observed that people who are engaged in social interactions tend to synchronize their body movements and language to match the behaviors of their conversation partners. This phenomenon, referred to as interpersonal coordination [1], has been identified in a wide variety of social contexts [2], including during digital interactions that are both asynchronous and text-based [3, 4]. It is believed that interpersonal coordination contributes to shared understanding in addition to increased feelings of empathy [5]. For example, [3] found that users of an anonymous peer support forum reported greater amounts of gratitude and emotional relief when the language of their peers was like their own. Interpersonal coordination plays a role in effective communication and the development of social relationships, with relevance to a variety of research areas (e.g., communication, human-computer interaction, marketing, psychology, and conversational agent development). Despite a long-standing emphasis on studying language, research can struggle to capture dynamic features that are core to coordination as a time-evolving phenomenon.

Language style matching (LSM) is an analytical tool that provides a window into interpersonal coordination by measuring the degree to which people mirror each other’s use of function words during conversation [6]. Function words are words that contribute to grammatical structure (e.g., articles, pronouns) as opposed to meaning [7], and describing interpersonal coordination according to function word frequency lends itself to comparisons across contexts. Although a variety of linguistic synchrony measures have been identified [8], LSM stands out because of its interpretability, predictive power, and ease of use [9]. Moreover, LSM is linked to empathy, rapport, social engagement, perceptions of support, and relationship stability across professional and romantic contexts [10–14]. Overall, LSM has improved our understanding of interpersonal coordination in various forms of media [15].

Although LSM is a successful measure of interpersonal coordination during conversation, it is limited in its ability to describe communication dynamics. That is, interpersonal communication is an inherently dynamic process in which participants continuously send and receive messages, leading to changes in expectations and behaviors during subsequent interactions that can be difficult to predict [16, 17]. Despite communication being a dynamic process, the typical approach for measuring LSM compares everything said by two speakers simultaneously without consideration of temporal order. This approach cannot describe how linguistic synchrony develops longitudinally nor in response to perturbations (e.g., a distracting or negative statement made by one of the participants). It is also reasonable to assume that the language of interlocutors will converge and diverge over time depending on the content of each participant’s message [18], and this dynamic information is lost through aggregation during standard LSM calculations. An understanding of language style dynamics could provide new insights on interpersonal coordination, but current methods do not account for the dynamic nature of communication.

The present work introduces a novel analysis for investigating longitudinal patterns in language style and LSM that is consistent with the conceptualization of communication as a dynamic process. This is achieved by combining the LSM equations with those from recurrence quantification analysis [RQA; 19]. RQA is a classical method

from nonlinear dynamical systems theory that is robust against outliers and makes minimal assumptions about the underlying data. In addition to turn-, speaker-, and conversation-level descriptions of LSM, our hybrid recurrence-based LSM (rbLSM) approach extracts features related to the stability and organization of communication, including estimates of determinism and laminarity. We begin by detailing how rbLSM can be applied to speech data collected from individuals and dyads. Then, we conduct a series of examples to demonstrate how the method can be used as a tool for both qualitative and quantitative analysis. Applied to monologue data from one speaker, our analysis models change in language style over the course of an orator’s speech (and text). Then, applied to facial EMG data, our method confirms significant linkages between dynamic LSM measures and the emotional qualities of conversations which were not detected by the traditional linear and static approach.

2 Related Work

2.1 Significance of Language Style Matching

The study of language style is motivated by the observation that there are substantial individual differences in how people form their sentences, and these stylistic differences are related to the speaker’s underlying psychological state [7]. Function words provide structure to sentences as opposed to content, thus measuring their frequency is an intuitive way of describing an individual’s language style [6, 7, 9]. Consistent with the idea that language style has implications for psychological state, function word frequency has been linked to cognitive processing and emotion [7]. For example, a meta-analysis of twenty-one papers using LIWC to measure the frequency of first-person singular pronouns reported a significant positive correlation between first-person pronoun usage and the severity of depression symptoms [20]. Notably, analysis of language style supports comparisons across contexts, since function words are relatively independent from topic, compared to content words.

LSM was proposed by [6] to describe linguistic synchrony in language style, as measured by the degree of similarity between two speakers’ function word frequency [21]. The function word categories typically included in LSM are articles, adverbs, auxiliary verbs, conjunctions, impersonal pronouns, prepositions, personal pronouns, and negations [9]. LSM scores range from zero (no similarity) to one (if the overall profile of function word usage is identical between texts). [22] suggest scores below .60 are indicative of low matching while scores above .85 are indicative of high matching, and these guidelines are generally consistent with the statistics reported by [9]. However, no validated cutoffs have been published in the literature. As with language style, the magnitude of LSM scores can be interpreted and compared in different contexts because it relies on function word frequency as opposed to content words.

Considering the connection between an individual’s language style, cognitive state, and social processing [7], it follows that LSM between conversation partners is related to the outcomes of their interactions. The tendency for speakers to match each other’s language style is believed to occur without conscience awareness, and it is particularly evident when speakers have close relationships [9]. For example, studies have reported an association between high LSM and post-interaction ratings of likeability

in small groups [23] and dyads [10, 24]. In contrast, there have been several studies on team coordination that have found negative relationships between team-wide LSM, task performance [25], and trust in teammates [26]. These findings suggest there are contexts, like teaming, in which ‘too much’ coordination and uniformity of behavior can be detrimental to performance [27]. Despite these conflicting results, the evidence is generally consistent with the idea that higher LSM improves the quality of social interactions.

In addition, LSM has been implicated in emotional processing. LSM varies depending on participants’ emotional state and conversation content [13, 28, 29]. [29] compared psychological survey scores for depression, generalized anxiety, and social anxiety to the amount of LSM in the text messages of 458 adults. The authors found that participants with self-report scores meeting clinical cutoffs for generalized and social anxiety had significantly lower text message LSM than those who did not. Similarly, [30] analyzed the text messages of young adults and found that those with elevated past year depression symptoms had increased amounts of LSM in their text messages. More broadly, increased LSM has been linked to the use of positive language during text-based negotiations [13], increased emotional reactivity in mother and child dyads [31, 32], increased empathetic communication by supervisors towards employees [10], and lower self-esteem scores [33]. Clearly, the relationship between LSM and emotion is complex, correlating with both positive and negative emotions depending on context. One explanation provided in the literature is that the automatic synchrony of function words increases when conversation participants are in a shared emotional state [9, 33].

2.2 Modifications to LSM

Given the diverse research areas interested in communication dynamics, LSM has been modified in several ways to accommodate different contexts of interest. One suggested modification by [21] involves averaging the pairwise LSM between every member of the group into a single estimate of synchrony. Alternatively, a one-to-many comparison can be made in which one speaker’s function word frequency is compared to the group’s average, reflecting that speaker’s conformity to the whole [21]. Both approaches would be appropriate for examining group-level interactions aggregated across entire conversations, and some variations on these calculations have been utilized previously [e.g., 14]. However, the staggering amount of aggregation required is likely to suppress any nuance related to intragroup relationships. Furthermore, none of these approaches address the lack of dynamical information provided by LSM.

An approach to looking at changes in LSM over time involves sequentially comparing each statement to the next in a conversation [6, 9] without accounting for the fact that conversations tend to be nonstationary and nonlinear. For example, reciprocal LSM is a recent modification that incorporates turn-level information into the final score and accommodates for group interactions [34]. Reciprocal LSM presumes an alternating turn-based conversation, such that speaker A makes a statement followed by speaker B, repeating until the end of the interaction. The relative similarity of function word categories is calculated for each statement in comparison to the previous and aggregated. The downside to these types of linear and sequential approaches is that

synchrony during interpersonal coordination frequently occurs at lags greater than one [35], meaning that a particular comment’s influence can reverberate throughout the conversation. There could be meaningful synchrony between temporally distant events even when there are insignificant amounts of synchrony between consecutive events. In other words, conversations are not linear and different statements have different impacts on participants. By only considering the similarity between consecutive statements, linear sequential approaches ignore more complex, and potentially informative, temporal patterns in language style that appear over the duration of an interaction.

2.3 Recurrence Quantification

Nonlinear dynamical systems theory forwards tools for understanding time-evolving behaviors [36], including descriptions of how systems coordinate and interact [37]. Among these tools, RQA has gained popularity in the social sciences because of its simplicity and flexibility in describing many types of systems [38]. The original procedure, auto-RQA, examines a single timeseries using self-similarity matrices to identify when that timeseries exhibits a repetitive behavior [38]. Multivariate extensions include multidimensional-RQA [39] for multiple variables, cross-RQA for comparisons between two systems [40, 41], joint-RQA for multiple systems with different measurement variables [19], and combinations of the former [e.g., 42]. Given the number of adaptations, it is appropriate to think of RQA as a family of related techniques that have evolved substantially over the past two decades. In the current paper, we focus on integrating LSM into cross-RQA for comparisons between dyads and auto-RQA for examinations of monologues.

The family of RQA techniques are united by a common set of procedures. RQA calculates the similarity between all possible data pairings within (or between time-series), and data pairs that are highly similar are flagged as recurrent. In the context of linguistic analysis, this means that every statement in a text passage is systematically compared to all other statements in that passage. The result is binary plot visualizing the similarity between every combination of text segments across all time lags, including within-subject comparisons when there is one speaker or between-subjects comparisons when there is multiple. The plot produced by RQA can be examined qualitatively or mined to quantitatively describe temporal patterns. The most common quantification in the literature is recurrence rate, which is essentially a nonlinear correlational estimate [19].

Although there is some practical overlap between correlation analyses and RQA, the RQA framework is particularly well-suited for describing the dynamics of interpersonal coordination. In addition to recurrence rate, it is possible to describe the temporal structure of interactions, including estimates of predictability (i.e., determinism), disorder (i.e., line entropy), and consistency (i.e., laminarity). Particularly relevant to understanding turn-level LSM dynamics, RQA identifies coordination that arises at time lags greater than one. As a result, a follow-up lag analysis can be applied to determine the delay required to maximize synchrony between participants and whether there is a leader-follower relationship. RQA is also appropriate for virtually any data type, including those that are nonlinear, have irregular sampling rates, or

unpredictable signal spikes. Taken together, RQA is a powerful framework for describing system dynamics that can be used to address the shortcomings of turn-level LSM analysis.

2.4 Previous applications of RQA to Linguistic Synchrony

Existing literature has applied cross-RQA to examine linguistic synchrony between two speakers, but these applications are distinct from the method proposed in the current. The most similar approach involves segmenting speech to individual words and categorizing them to a binary value indicating whether each is a function word. In this way, categorical RQA is used to examine repetitive usages of individual function words and the change in word choice over time [34, 43]. In contrast, the current paper presents a turn-level analysis. Our approach directly integrates the LSM equation into RQA instead of applying categorical transformations. In doing so, it is more consistent with the original conceptualization of language style and LSM. Other RQA-based methods for measuring linguistic synchrony include categorizing speech into binary variables [44, 45], comparing utterances directly [46, 47], coding text sequences according to content [48], and analyzing similarity in audio qualities [e.g., speech rate, volume; 49]. Overall, we consider our approach to be distinct and oriented towards answering research questions specific to turn- and speaker-level language style dynamics.

2.5 rbLSM: Hybridizing Recurrence and Language Style Matching Analyses

LSM is tailored to the study of language; RQA is used to uncover dynamic values and patterns defining of a system. Their operations provide a point of intersection for the two analyses, such that the proposed rbLSM quantifies dynamic features of LSM. The current work integrates LSM into the RQA framework by replacing the Euclidean distance formula with LSM as the similarity metric. As a result, two separate matrices are generated that can be used for linguistic analysis. The first is the LSM matrix, which is a similarity matrix containing the typical LSM scores (ranged 0 to 1) in each cell reflecting the degree of similarity between statement pairs in a conversation or text. Turn-, speaker-, or conversation-level LSM measures can be mined from this matrix. For example, the similarity of consecutive statements can be calculated by averaging scores along the primary diagonal, resulting in scores like reciprocal LSM [34]. The second matrix is the recurrence plot, which flags statement pairs as having high or low LSM depending on a predetermined threshold. Like the LSM matrix, dynamic information from the recurrence plot can be retrieved using any classical recurrence quantification, including recurrence rate, determinism, entropy, or laminarity [19]. In what follows, we provide a detailed description of rbLSM. In addition, we discuss how LSM can be integrated with multiple types of RQA to accommodate for different social context (i.e., one-speaker, dyads, groups). Then, we demonstrate the technique to analyze language in a monologue and between dyads.

3 Method

3.1 Calculating rbLSM

The first step is generating a similarity matrix that compares each statement made by Speaker A to statements made by Speaker B. It is assumed that the conversation is segmented by turn, with each turn alternating chronologically beginning with Speaker A. If analyzing the speech of an individual speaker, then the data series is compared to a copy of itself (i.e., auto-RQA). There is limited discussion about the word count required for the LSM. [50] recommends a minimum word count of 25 to 50 when investigating psychological text features with word frequencies. However, a similar justification is not given in the case of LSM, and we present our example analysis without a minimum. Before LSM can be calculated, it is necessary to count the relative frequency of articles, adverbs, auxiliary verbs, conjunctions, impersonal pronouns, prepositions, personal pronouns, and negations appearing in each statement.

Formulas for calculating LSM between two statements are provided in Eq. 1 and 2. The relative frequency of a function word category in a statement is described as $category_{A(i)}$ or $category_{B(j)}$, with i and j corresponding to the chronological order of the statement. For example, $Adverb_{B(5)}$ would refer to the relative frequency of adverbs in Speaker B’s fifth statement. The similarity between two statements, $LSM_{category(i,j)}$, is calculated separately for each category as the relative difference between $category_{A(i)}$ and $category_{B(j)}$ (Eq. 1). Then, the overall LSM for a pair of statements (i,j) is taken as the average similarity across all categories (Eq. 2). As a plain language example, if Speaker A’s statement contained 4% adverbs compared to Speaker B’s statement with 5% adverbs, the category-level similarity (Eq. 1) between those statements would be approximately 89%. That category-level similarity is calculated for each function word type and averaged to return the LSM (Eq. 2).

$$LSM_{category(i,j)} = 1 - \frac{|category_{A(i)} - category_{B(j)}|}{category_{A(i)} + category_{B(j)} + .0001} \quad (1)$$

$$LSM(i,j) = average(LSM_{adverb(i,j)} + LSM_{article(i,j)} + \dots) \quad (2)$$

The LSM matrix, Eq. 3, is generated by calculating LSM between every pair of Speaker A’s statements, i , and Speaker B’s statements, j . Thus, the score contained in coordinates $(1, 3)$ of the LSM matrix would correspond to the LSM between speaker A’s first statement and Speaker B’s third statement. If there was a single speaker, it would reflect the self-similarity between that speaker’s first and third statement because the analysis would compare that speaker’s data to itself.

$$LSM_{matrix_{i,j}} = LSM(i,j) \quad (3)$$

The LSM matrix is transformed into a recurrence plot (RP) using a Heaviside function (Eq. 4) that converts scores above a critical threshold, r , to 1’s (matching)

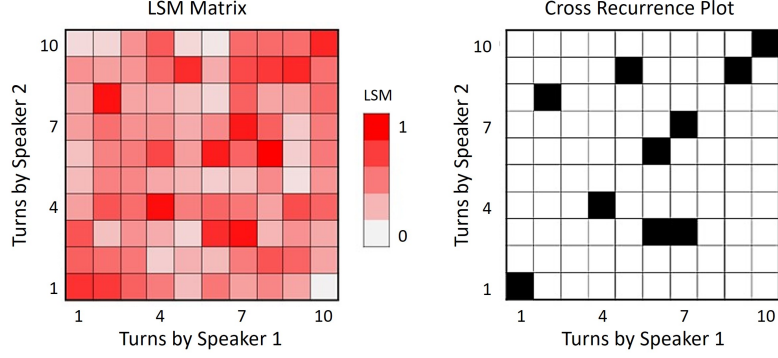


Fig. 1 Example LSM matrix (left) and RP (right) using cross-RQA. The LSM matrix is color coded according to LSM values from 0 (no overlap) to 1 (complete match). Statement pairs with high degree of matching (LSM > 0.75) are flagged by the black tiles in the RP.

and others to 0's (non-matching). The value of r should be held constant throughout the sample to facilitate statistical comparisons. We include the equation for recurrence rate (RR; Eq. 5), which reflects the percentage of the RP with highly synchronized language style. An example LSM matrix and RP are provided in Figure 1. When there is one speaker, the RP makes within-subject comparisons of all that speaker's statements. When there are two speakers, the RP makes between-subjects comparisons from all pairings of their statements.

$$RP_{i,j} = \begin{cases} 1 & \text{if } LSM_{matrix_{i,j}} \geq r \\ 0 & \text{elsewhere} \end{cases} \quad (4)$$

$$RR = \frac{\sum_{i=1}^n \sum_{j=1}^n RP_{i,j}}{n^2} \quad (5)$$

Special attention should be paid to the threshold, r , because it is tied to the RR. Traditionally, r is a somewhat arbitrary value that varies by dataset. In the case of rbLSM, the value of r is not arbitrary because LSM scores are relative absolute differences that can be compared across contexts. The only guidance in the literature concerning what is high or low LSM is provided by [22], who suggests that high LSM values are those above 0.85. Anecdotally we observed that r values between 0.75 and 0.85 yielded RRs in the range of 5-10%, consistent with classical rule-of-thumb approaches. Therefore, a value of 0.75 can serve as a starting point, with the caveat that r may need to be adjusted if the sample average RR is high or low enough to cause floor or ceiling effects in the sample distribution.

3.1.1 Understanding and Mining rbLSM Matrices

To understand how the matrices in Figure 1 can be mined for dynamical information, it is necessary to describe their structure. Column and row numbers correspond to

Table 1 Levels of rbLSM aggregation

Outcome	Aggregation	Interpretation
Turn-level similarity	Column wise for speaker A or row wise for speaker B. Either used for monologues	Similarity between a statement and partner’s overall language style
Turn-level Reciprocity	Column wise in lower triangle exclusive diagonal for Speaker A or row wise in upper triangle inclusive diagonal for Speaker B.	Similarity between one statement and language style in partner’s preceding statements
Speaker-level Reciprocity	Lower triangle exclusive of diagonal for Speaker A or upper triangle inclusive for Speaker B	Overall degree to which one speaker’s language style match partner’s preceding language
Overall Scores	Matrix-wide aggregation	Conversation-level similarity

specific statements made by Speaker A and B. That is, the n th column compares Speaker A’s n th statement to Speaker B’s statements. Likewise, the n th row compares Speaker B’s n th statement to Speaker A’s statements. The primary diagonal compares statements at a lag of zero (i.e., $i = j$ in Eq. 3). With Speaker A making the first statement in the conversation, this means that the primary diagonal compares Speaker A’s statements with the immediate reply given by Speaker B. The diagonal below the primary (i.e., $i + 1 = j$) includes comparisons between all of Speaker B’s statements with the immediate reply by Speaker A. Together, these two diagonals contain all consecutive statement pairings analyzed by the reciprocal LSM method proposed by [34]. Diagonals further from the primary make comparisons at increasing time lags. Extending this logic, cells in the upper triangle inclusive of the primary diagonal (i.e., $i \leq j$) pair Speaker B’s statements with prior statements by Speaker A. The inverse is true for the lower triangle excluding the primary diagonal (i.e., $i > j$). Thus, the upper and lower triangle are reflective of each speaker’s overall language style reciprocity, or the tendency to mirror the language used by their partner leading up to their own statements. This logic does not extend to the analysis of individual speakers where the diagonal instead reflects the similarity between one speaker’s statement with itself and the upper and lower triangles are symmetrical.

The LSM matrix and RP can be aggregated to turn-, speaker-, or conversation-level insights depending on research question (Table 1). The LSM matrix (left, Figure 1) is aggregated through averaging, while the RP (right, Figure 1) is aggregated according to the RR within a region of the matrix. Utilizing LSM versus RP has implications for how the output is interpreted. That is, averages from the LSM matrix reflect central tendency, whereas the RR reflects the frequency of highly elevated LSM. It is possible for dyads to speak with high LSM on average despite only a few of their interactions meeting the threshold, r , required for a recurrence in the RP.

The list of outcomes provided in Table 1 is not exhaustive. We do not include calculations for several dynamic measures from cross-RQA or auto-RQA, including

determinism, line entropy, and laminarity. Descriptions for those measures are available in [19, 40]. Diagonal cross-recurrence profiles can also be used with rbLSM to identify leader-follower dynamics [40].

4 Analysis of Language Style Dynamics in Monologues

4.1 Procedure

In what follows, we apply rbLSM with auto-RQA to describe language style in a speech given by Carl Sagan to demonstrate rbLSM in the context of monologues. In specific, we analyze a passage from the 1994 book “Pale Blue Dot” provided by the Planetary Society [51]. This text was selected because of the sentence variety and organization, with sentences ranging from three to eighty words in length. The text was segmented by sentence to form a chronological data series. The LIWC 2022 software package [52] was used to extract the frequency of each function word category per sentence. After processing, there was a total of 20 sentences with an average word length of 18.15 ($SD = 18.14$) and 60.85% average function words ($SD = 21.39\%$). A radius of .55 was included in the calculation of the RP.

To provide a comparison point, we also apply rbLSM to analyze the first 20 sentences from Chapter I of *On the Origin of Species* [53]. This text has entered the public domain, and the portion of text analyzed here is provided in the Appendix. Both passages concern scientific topics (i.e., astronomy and biology). However, Sagan’s speech can be described as persuasive and philosophical, whereas Darwin’s work is a highly descriptive scientific essay about the theory of evolution. Therefore, this second example allows for a comparison between distinct types of communication and rhetorical motivations. After processing, there was a total of 20 sentences with an average length of 40.75 words ($SD = 20.98$) and 57.20% average function words ($SD = 7.47\%$). The same radius of .55 was applied.

4.2 Findings

The LSM matrix and RP for Sagan’s Pale Blue Dot speech can be found in Figure 2. Several patterns in language style dynamics are made visually apparent by the RP. Following along the primary diagonal from the lower-left to upper-right corner, there are several clusters of sentences with distinct language styles. For example, there is a high degree of similarity among the first five sentences. There is also a repeating pattern of language style among sentences 8 to 14 that recurs during sentences 17 to 20 (see Table 2). Although we are unable to share the transcript, this clustering aligns with the organization of the speech, which begins with a series of short sentences composed of pronouns, auxiliary verbs, and adverbs (i.e., “that’s here. that’s home.”; Table 2). The language style pattern in the first five sentences is repeated later in the speech between sentences 15 and 16 when the author shifts back to this abbreviated style for emphasis.

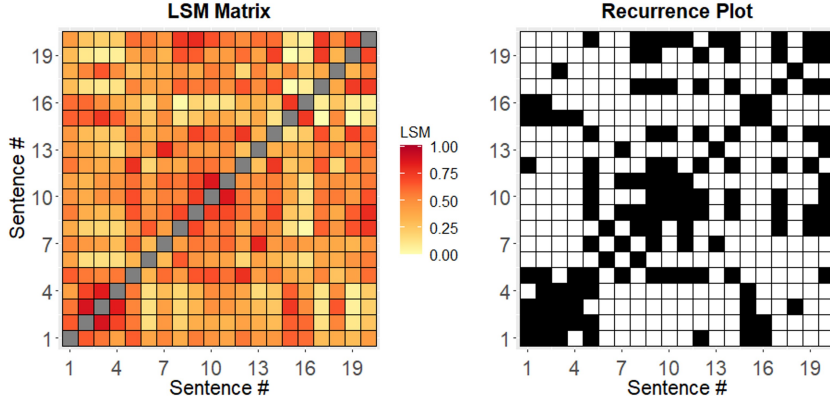


Fig. 2 LSM distance matrix (left) and recurrence plot (right) from Carl Sagan’s text. Values in the LSM matrix range from 0 (no overlap) to 1 (perfect match), whereas values in the recurrence plot are represented by white (non-recurrent language style) and black (highly similar language style). Diagonals in both matrices are ignored, and upper and lower triangles are symmetrical when analyzing speech from a single individual.

Table 2 Average language style of recurrent patterns identified in Pale Blue Dot

Function Word Category	Sentences 1-5	Sentences 8-14	Sentences 17-20
Articles	0%	6.69%	6.95%
Adverbs	15.74%	10.49%	10.55%
Auxiliary Verbs	31.74%	13.66%	11.03%
Conjunctions	0%	2.38%	2.97%
Negations	0%	2.94%	2.94%
Impersonal Pronouns	37.48%	13.91%	11.28%
Personal Pronouns	14.35%	8.01%	8.14%
Prepositions	6.61%	10.67%	11.33%

When using rbLSM to examine one speaker, matrix-wide LSM and RR reflect the self-similarity of speech over time. In other words, higher average LSM and RR correspond to a more consistent style of language across all speech segments. Either matrix can be used, with matrix-wide LSM providing a more intuitive description and RR being less sensitive to outliers. In this case, the average LSM between sentences was .43 ($SD = .19$) with a RR of 26.50%. The sentence that was most consistent with the overall language style of the speech was sentence 10, while the most distinct was sentence 16. Diagonal lines in the matrices indicate that the progression of language style observed in one consecutive sequence of segments was repeated elsewhere. The average diagonal line length in this speech was 3.13 and 47.17% of all recurrent points fell on a diagonal line (i.e., the determinism). Vertical lines represent consecutive statements in which every statement is identical to the others in the sequence. The average vertical line length in this speech was 2.20 and 51.89% of all recurrent points fell on a vertical (i.e., the laminarity). Together, these four line-based metrics are ways of describing language style stability and transitions between style ‘states.’ In comparison, the writing of Charles Darwin contained a more consistent language style with

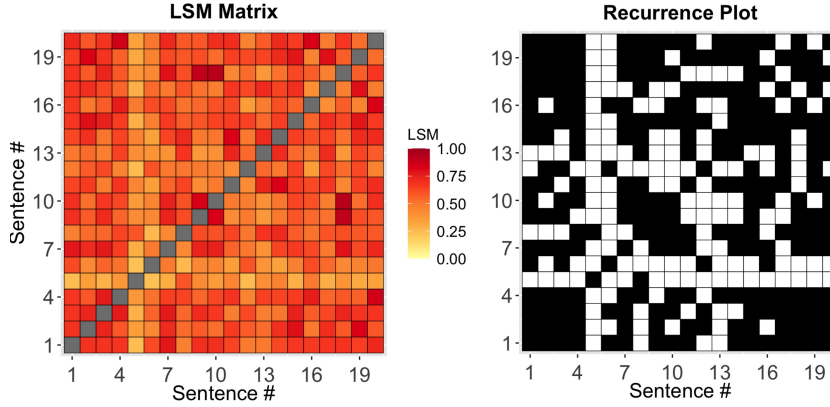


Fig. 3 LSM distance matrix (left) and recurrence plot (right) from Charles Darwin’s text.

an average LSM of .58 ($SD = .14$) and RR of 55.50% (see Figure 3). The percentage of recurrent points falling on a diagonal line was 83.78% with a laminarity of 84.23%. This indicates that, in addition to having consistent overall language style, the author tended to use the same language style in longer consecutive sequences and without major shifts or alternating patterns (as seen in the previous example). These findings are consistent with expectations of persuasive versus informative writing. That is, persuasive writing can use shifts in sentence structure, length, and tone for rhetorical impact [54], whereas informative writing tends to describe a topic more plainly. As such, research has shown that effective persuasive writing is associated with syntactic complexity and lexical diversity [55, 56].

5 Comparison between LSM Dynamics and Emotion in Dyads

5.1 Procedure

Next, we demonstrate how rbLSM can be applied for between-speaker comparisons with cross-RQA and test specific hypotheses related to interpersonal coordination. Specifically, rbSLM was used to analyze the emotional responses of dyads during unstructured conversation. Data was retrieved from a study on the effects of video co-viewing on communication [57]. Participants were undergraduate college students from Indiana University Bloomington, and the study protocol was approved by the University’s Institutional Review Board. We analyzed data from a subset of 65 dyads who completed a baseline task in which they were instructed to speak freely for two minutes while connected to physiological monitoring equipment. The dyads were assigned to complete this task in one of three experimental conditions: Audio conferencing, in-person communication, and video conferencing. Participants were seated next to each other for the in-person condition, whereas they communicated with their partner using a laptop for the audio and video conferencing conditions.

We analyzed data from participants’ audio and facial EMG recordings. Conversation audio was manually transcribed by the research team such that the resulting text was segmented by turn with timestamps for the start and end of each turn. Function word frequency was then extracted from each segment using the procedure described above. Facial EMG was recorded from the corrugator supercilia and orbicularis oculi muscles on the left side of the face using reusable 4mm AG-AGCL electrodes with isotonic gel. A comb band stop filter was used to remove 60hz interference and a 28-500 Hz bandpass filter was applied [57]. Finally, the processed data was down sampled to 1hz and baseline corrected. Activation of the corrugator supercilia muscle has been associated with negative emotion [58], whereas activation of the orbicularis oculi muscle has been associated with expressions of joy [59].

Elevated LSM is thought to contribute to prosocial interactions and emotion [13, 29]; therefore, the objective of this analysis was to examine the relationship between LSM dynamics and the emotional response of conversation partners. RbLSM was calculated with a radius of .50 to determine recurrent points. The radius was initially set to .75, but that value led to floor effects in the distribution of sample RRs. The radius was iteratively lowered until a full range of similarity scores could be represented. Two mixed-effects models were fit using the LME4 R software package [60] to compare the turn-level RR of a statement made by one speaker to the facial muscle activity of their partner when they listened to that statement. This approach allowed us to consider repeated-measures second-by-second EMG recordings nested at the turn and speaker levels. Condition, gender of the speaker, time, and word count were all included as control variables. Speaker and turn identifiers were included as random effects. The inclusion of word count as a control variable was to ensure that any potential differences in facial EMG predicted by RR were not driven by differences in statement length. It was predicted that higher RR would be associated with increased emotional reaction, measured by increased activation of the corrugator supercilia and orbicularis oculi muscles.

5.2 Findings

5.2.1 Descriptive Statistics

The distribution of turn-level LSM ($M = 0.49$, $SD = 0.09$) and RR ($M = .52$, $SD = 0.21$) scores are depicted in Figure 4. The number of statements made by individual speakers ranged from 5 to 19 ($M = 12.53$, $SD = 3.41$). The average diagonal line length was 2.83 ($SD = 0.48$) with a percentage determinism of 67.29 % ($SD = 11.81$). Likewise, the average vertical line length was 3.37 ($SD = 0.81$) with a laminarity of 74.00 % ($SD = 13.60$). A scatter plot is provided in Figure 5 that compares the degree each speaker reciprocated the language style of all preceding statements by their partner (i.e., representing linguistic accommodation). Points in Figure 5 further from the diagonal indicate that one member of the pair was more accommodating than the other. An example RP of dyads with low versus high linguistic coordination is provided in Figure 6.

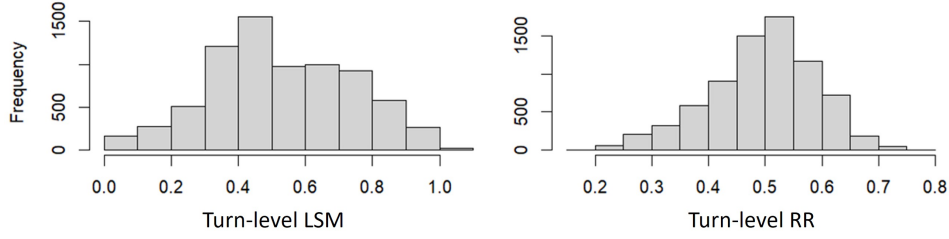


Fig. 4 Distribution of turn-level LSM and RR scores. Higher values reflect increased linguistic synchrony.

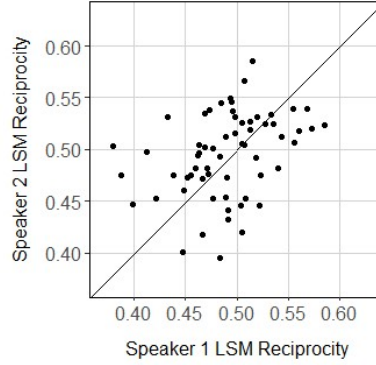


Fig. 5 Comparison of speaker-level rbLSM scores. Points that are further from the diagonal represent an imbalance between how much one speaker reciprocated the language style of their partner.

5.2.2 Facial EMG Analysis

Analysis of dyads' facial EMG (Table 3) revealed the degree of turn-level RR in spoken statements predicted increased activity of observer orbicularis oculi muscle ($\beta = 0.22$, $p = .006$) but not the corrugator supercilii ($p = .817$). This effect was observed when controlling differences in word count ($ps = .501, .591$). There was also a negative longitudinal trend in observer orbicularis oculi ($p = .006$). Given the orbicularis oculi muscle is linked to expressions of joy, these results indicate that turn-level linguistic coordination is associated with joyful muscle movements in the face of the person who is listening to that statement. These findings demonstrate that individual differences in LSM dynamics are predictive of communication outcomes, such as the second-by-second emotional processing of conversation participants while listening to a linguistically accommodating message. Moreover, this analysis was achieved with a granularity that would not be possible with traditional approaches to LSM.

5.2.3 Comparison to linear and sequential turn-level approaches

One motivation for using rbLSM versus manually calculating LSM in sequential statement-reply chains is that the latter approach assumes that (1) the language style

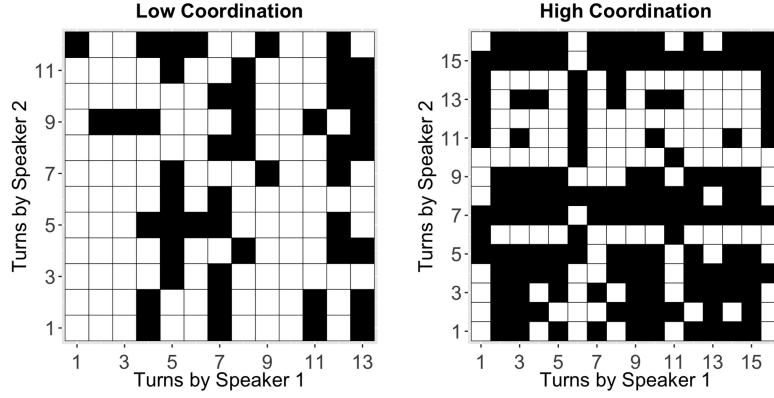


Fig. 6 Example of two dyads representing low (left) and high language coordination. The RR of the low-coordination pair was 31%, and the RR of the high-coordination pair was 56%.

Table 3 Mixed-effects model predicting facial EMG from turn-level language dynamics

<i>Predictors</i>	Observer Corrugator Supercilii			Observer Orbicularis Oculi		
	<i>Estimates</i>	<i>CI</i>	<i>p</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	-0.15	-0.39 – 0.10	.236	-0.12	-0.44 – 0.20	.462
[Condition] In-person	0.29	0.06 – 0.53	.015	0.31	-0.01 – 0.63	.057
[Condition] Video	0.13	-0.11 – 0.37	.300	0.14	-0.18 – 0.47	.387
[Gender] Female	-0.00	-0.22 – 0.21	.963	0.11	-0.17 – 0.40	.437
[Gender] Unreported	0.09	-0.23 – 0.42	.579	-0.09	-0.53 – 0.34	.676
Turn-level RR	0.02	-0.12 – 0.15	.817	0.22	0.06 – 0.37	.006
Time	-0.00	-0.00 – 0.00	.360	-0.00	-0.00 – -0.00	<.001
Word Count	0.00	-0.00 – 0.00	.591	0.00	-0.00 – 0.00	.501
Random Effects						
σ^2	0.59			0.32		
τ_{00} (Dyad)	0.30			0.53		
τ_{00} (Turn)	0.09			0.22		
ICC	0.38			0.69		
N (Dyad)	123			123		
N (Turn)	1543			1543		
Observations	7279			7279		
Marginal R^2	0.015			0.037		
Conditional R^2	0.392			0.706		

of one statement is entirely dependent on the previous and (2) does not account for synchronous patterns appearing in time periods larger than a lag of one. In other words, rbLSM can capture nonlinear and nonstationary patterns in language style that a sequential approach cannot. To exemplify this, we compared turn-level RR from rbLSM to a turn-level LSM scores calculated by sequentially comparing each statement to the next. There was a weak correlation between turn-level RR and these sequential LSM scores ($r = .34$, $p < .001$), suggesting related but distinct constructs. Finally, we

examined the relative predictive power of a linear sequential approach by testing it in the model described in the previous section. However, unlike the recurrence-based estimation of LSM, sequential LSM was not a significant predictor of observer corrugator supercilia ($\beta = 0.03$, $p = .66$) nor orbicularis oculi ($\beta = 0.10$, $p = .23$) activity. Taken together, these findings provide evidence that our nonlinear approach for measuring LSM dynamics offers unique insights above and beyond that of a sequential approach, which can be attributed to RQA’s sensitivity to nonlinear and nonstationary language patterns appearing intermittently over time.

6 Discussion

The aim of this work was to develop a technique for describing LSM dynamically, providing inroads for extracting new linguistic dynamic features. We combine advantages of LSM in being tailored to identifying linguistic features with RQA’s theoretical foundation for understanding the dynamics of interpersonal coordination. In the sections above, we provided a detailed description of how rbLSM is calculated. Then we demonstrated the utility of rbLSM in a series of examples that explore language style through visual examination of rbLSM recurrence plots and quantification of language style stability and synchrony. Specifically, we show how rbLSM can provide descriptions of the structure, stability, and temporal evolution of both single speaker monologues and dyadic conversation that would not be possible using a standard linear approach, and we uncover linkages between our nonlinear rbLSM measures to second-by-second emotional processing. Overall, rbLSM provides a means for investigating dynamics of language style across varied interpersonal contexts.

The example analysis of Sagan’s 1984 speech [51] highlights how rbLSM can be used to describe the stability and organization of speech based on vertical and diagonal line structures in auto-recurrence plots. In comparison to a strictly explanatory text, Sagan’s persuasive and philosophical speech was more varied and contained an alternating language style for rhetorical impact. We analyzed these two passages to highlight the ability of rbLSM to identify temporal language style patterns and differentiate between writing genres. In practice, more data points would be needed to make statistical inference, but the ability to quantify both local fluctuations and longitudinal trends in an individual’s speech points to a variety of future applications [61]. For example, a common goal of dynamical systems research is to understand how systems respond to perturbations. In the context of communication, a perturbation could involve an event that disrupts a speaker’s cognitive processing. Thus, rbLSM could be implemented to compare the speech dynamics of participant groups (e.g., of varying anxiety levels) when perturbations are introduced to impromptu speech to determine how language style responds to acute stress or emotion. Similar investigations would not be possible using static implementations of LSM because they do not consider time dependence.

The example analysis of facial EMG provides a more extensive investigation into the relationship between LSM dynamics and emotional processing. Our findings are consistent with previous research linking interpersonal coordination to social connectedness [62] and positive affect [62, 63]. Research on static LSM measures have

reported similar linkages with positive emotions, such as reporting positive relationships between LSM, engagement, and liking [13, 24]. Our findings differ from these prior works with respect to the granularity achieved through rbLSM, which enabled direct comparisons between segments of verbal turn taking and nested 1 hz physiological recordings. In fact, the method we used to compare turn-level LSM to physiological response is one potential route for investigating the hypothesis that LSM increases when people are in a shared, or intense, emotional state [9, 33]. Our results provide convergent evidence that rbLSM can capture nuanced differences in turn-level language style with appreciable associations to interpersonal outcomes such as onlooker emotion.

The use of RQA to examine language style aligns with a dynamic systems approach to investigating communication [61, 64–67]. Interpersonal communication is a dynamic process in which participants reciprocally influence one another, and the qualities of their communication evolve in response to feedback from interlocutors and the environment [61, 66]. Communication occurs in a hierarchy of nested interpersonal and intrapersonal processes, including the cognitive and physiological subsystems that underly the behavior of interlocutors [66]. Moreover, statements expressed during one turn in a conversation depend on the verbal communication exchanged in previous turns, with the ensemble of these interactions leading to emergent patterns in language [67]. Our analysis addresses fundamental topics in the dynamical study of communication. Measuring recurrence rate, determinism, and laminarity provides information about system stability and transitions between linguistic ‘states.’ The fact the analysis can capture nonlinear and nonstationary patterns means that it is possible to compare local fluctuations in language style to outcomes that are unfolding at larger or smaller timescales [61]. Finally, configuring the analysis for dyads reflects how communication systems transfer information and entrain to one another. Taken together, rbLSM provides more detailed information about language style dynamics than a simple pairwise comparison of turn taking.

In addition to analyzing in-person communication, rbLSM is a strong candidate for analyzing discrete interactions in digital mediums. Recent interest in a dynamic systems approach is partially spurred by the availability of high-resolution time series data [61]. Social media data makes it possible to model interpersonal communication over the course of years and decades. However, most social media interactions are asynchronous and discrete, leading to unique communication qualities [68]. RbLSM is well-situated in this respect because it does not consider the temporal proximity of statements, and it models turn taking as a discrete process. Notably, the analysis can be reconfigured depending on the predominant form of communication on the platform (e.g., direct messaging, messaging boards). Previous projects have applied mdRQA to model the stability of low-level comment features (e.g., word count) on social network sites [4], but more work is needed to determine if rbLSM would perform similarly well in digital contexts.

Finally, it is worth discussing the number of ways rbLSM can be modified to fit various experimental setups. The most obvious modification is changing the underlying recurrence analysis to match the number of people involved in an interaction. We demonstrated how auto-RQA can be used to measure language style dynamics within

one person’s speech and cross-RQA can be used to explore dyadic interactions. However, there are several options for scaling to groups. A group could be treated as a single system and analyzed using auto-RQA, capturing holistic changes in group language style. Alternatively, one member of a group can be compared to other interlocutors in a cross-RQA to isolate their individual involvement. It would even be possible to implement group LSM into multidimensional RQA [39] or joint RQA [19], albeit these implementations would not be as straightforward as the prior. Beyond modifying the underlying RQA, the parameters of the similarity matrix can be altered to fit other synchrony estimates [e.g., semantic cosine similarity; 69]. For example, the function word categories in the LSM equation could be replaced by dimensions of affect measured according to the frequency of sadness, fear, and anxiety related words. Drastic modifications using affective synchrony or lexical similarity instead of LSM may lead to confusion about the naming of rbLSM, but the point remains that RQA is sufficiently flexible to accommodate a variety of ‘language matching’ research questions without dramatically altering the interpretation.

6.1 Limitations

The technical outline of rbLSM included in the current paper leaves several analytical decisions up to the discretion of the researcher. Specifically, we do not provide guidance on whether to use average LSM scores from the distance matrices or recurrence rate from the recurrence plots when aggregating cell-wise values because there are merits to both. Recurrence rate is the most widely researched outcome of RQA and it is less likely to be biased when the data is skewed or nonlinear. On the other hand, LSM scores are non-trivial values that can be compared across studies, which is pragmatically valuable. We expect the two metrics will be interchangeable in many cases. For example, the Pearson correlation between turn-level LSM and RR in the facial EMG example analysis was .87 ($p < .001$).

In addition, research is needed to examine the utility of specific RQA outcomes. The current facial EMG analysis is focused on comparing local fluctuations in speaker LSM to onlooker emotion but does not consider organizational outcomes such as determinism, vertical line distributions, or specific lag descriptions. Turn-level recurrence rate is implemented for this analysis because the research question concerned momentary emotional reaction, and the metrics derived from diagonal or vertical line structures unfold at longer time scales than applicable. The most appropriate comparison with the available data was using scores on the Ten Item Personality Inventory [70]. A follow-up correlation analysis revealed there was a significant negative correlation between determinism and participant extroversion, $r(63) = -.20$, $p < .05$. This suggests that extroverted dyads had language style patterns that were more difficult to predict. However, these findings are exploratory, and future research should continue to examine conversation-level outcomes.

7 Conclusion

This study introduced a novel technique for investigating language style dynamics, which we refer to as recurrence-based LSM. Unlike related methods that are focused on

continuous speech features (e.g., pitch) or categorization of utterances, rbLSM models turn-level interactions by integrating the LSM similarity metric into the recurrence quantification framework. This hybrid approach makes it possible to describe language style dynamics according to both standard LSM scores and recurrence-based quantifications like recurrence rate, determinism, laminarity, entropy, and leader-follower estimates. To demonstrate its utility, the method was applied to examine language style dynamics in both individual speech (auto-RQA) and between dyads (cross-RQA). Our results suggest that rbLSM estimates of language style dynamics are predictive of positive facial expressions in dyads. Beyond these example applications, rbLSM is particularly well-suited for describing asynchronous social interactions on social media. Future directions include expanding the analysis to interactions between large groups (e.g., using mdRQA) and exploring how the technique can be altered to accommodate different aspects of language matching.

Statements and Declarations

Funding No funds, grants, or other support was received.

Competing interests The authors have no relevant financial or non-financial interests to disclose.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Data availability Data from the monologue analysis is publicly accessible via <https://www.planetary.org/worlds/pale-blue-dot> and in the Appendix.

Code availability Not applicable.

References

- [1] Cornejo, C., Cuadros, Z., Morales, R. & Paredes, J. Interpersonal coordination: methods, achievements, and challenges. *Frontiers in psychology* **8**, 1685 (2017).
- [2] Wiltshire, T. J., Philipsen, J. S., Trasmundi, S. B., Jensen, T. W. & Steffensen, S. V. Interpersonal coordination dynamics in psychotherapy: A systematic review. *Cognitive Therapy and Research* **44**, 752–773 (2020).
- [3] Doré, B. P. & Morris, R. R. Linguistic synchrony predicts the immediate and lasting impact of text-based emotional support. *Psychological Science* **29**, 1716–1723 (2018).
- [4] Necaise, A., Williams, A., Vrzakova, H. & Amon, M. J. *Regularity versus novelty of users’ multimodal comment patterns and dynamics as markers of social media radicalization*, 237–243 (2021).
- [5] Wood, A., Lipson, J., Zhao, O. & Niedenthal, P. in *Forms and functions of affective synchrony* 381–402 (Springer, 2021).

- [6] Niederhoffer, K. G. & Pennebaker, J. W. Linguistic style matching in social interaction. *Journal of language and social psychology* **21**, 337–360 (2002).
- [7] Chung, C. & Pennebaker, J. in *The psychological functions of function words* 343–359 (Psychology Press, 2011).
- [8] Church, R. B., Kelly, S. & Holcombe, D. Temporal synchrony between speech, action and gesture during language production. *Language, Cognition and Neuroscience* **29**, 345–354 (2014).
- [9] Ireland, M. E. & Pennebaker, J. W. Language style matching in writing: synchrony in essays, correspondence, and poetry. *Journal of personality and social psychology* **99**, 549 (2010).
- [10] Meinecke, A. L. & Kauffeld, S. Engaging the hearts and minds of followers: Leader empathy and language style matching during appraisal interviews. *Journal of Business and Psychology* **34**, 485–501 (2019).
- [11] Ireland, M. E. *et al.* Language style matching predicts relationship initiation and stability. *Psychological science* **22**, 39–44 (2011).
- [12] Sanova, A. *Beyond Words: Language Style Matching as an Index of Therapeutic Alliance in Psychotherapy*. Master’s thesis, University of California, Irvine (2022).
- [13] Ireland, M. E. & Henderson, M. D. Language style matching, engagement, and impasse in negotiations. *Negotiation and conflict management research* **7**, 1–16 (2014).
- [14] Rains, S. A. Language style matching as a predictor of perceived social support in computer-mediated interaction among individuals coping with illness. *Communication Research* **43**, 694–712 (2016).
- [15] Chen, L., Baird, A. & Straub, D. A linguistic signaling model of social support exchange in online health communities. *Decision Support Systems* **130**, 113233 (2020).
- [16] Fogel, A. Dynamic systems research on interindividual communication: The transformation of meaning-making. *Journal of developmental processes* **1**, 7–30 (2006).
- [17] Foulger, D. Models of the communication process. *Brooklyn, New Jersey* 1–13 (2004).
- [18] Gallois, C., Ogay, T. & Giles, H. Communication accommodation theory. *Theorizing about intercultural communication* 121–148 (2005).
- [19] Marwan, N., Romano, M. C., Thiel, M. & Kurths, J. Recurrence plots for the analysis of complex systems. *Physics reports* **438**, 237–329 (2007).

- [20] Edwards, T. & Holtzman, N. S. A meta-analysis of correlations between depression and first person singular pronoun use. *Journal of Research in Personality* **68**, 63–68 (2017).
- [21] Boyd, R. L., Ashokkumar, A., Seraj, S. & Pennebaker, J. W. The development and psychometric properties of liwc-22. *Austin, TX: University of Texas at Austin* **10**, 6 (2022).
- [22] Cannava, K. & Bodie, G. D. Language use and style matching in supportive conversations between strangers and friends. *Journal of Social and Personal Relationships* **34**, 467–485 (2017).
- [23] Gonzales, A. L., Hancock, J. T. & Pennebaker, J. W. Language style matching as a predictor of social dynamics in small groups. *Communication Research* **37**, 3–19 (2010).
- [24] Ireland, M. E. *Three explanations for the link between language style matching and liking*. Ph.D. thesis, The University of Texas Austin (2011).
- [25] Heuer, K., Müller-Frommeyer, L. C. & Kauffeld, S. Language matters: The double-edged role of linguistic style matching in work groups. *Small Group Research* **51**, 208–228 (2020).
- [26] Carmody, P. C., Mateo, J. C., Bowers, D. & McCloskey, M. J. *Linguistic coordination as an unobtrusive, dynamic indicator of rapport, prosocial team processes, and performance in team communication*, Vol. 61, 140–144 (SAGE Publications Sage CA: Los Angeles, CA, 2017).
- [27] Amon, M. J., Vrzakova, H. & D’Mello, S. K. Beyond dyadic coordination: Multimodal behavioral irregularity in triads predicts facets of collaborative problem solving. *Cognitive science* **43**, e12787 (2019).
- [28] Schaper, R., Nowotny, C., Michalek, S., Schmidt, U. & Brockmeyer, T. Language style matching and treatment outcome in anorexia nervosa. *European Eating Disorders Review* **31**, 110–120 (2023).
- [29] Stamatis, C. A. *et al.* The association of language style matching in text messages with mood and anxiety symptoms. *Procedia computer science* **206**, 151–161 (2022).
- [30] Weinstein, M. & Jensen, M. Depression and reciprocal language style matching in text messages. *Journal of Language and Social Psychology* **43**, 37–62 (2024).
- [31] Borelli, J. L. *et al.* Language matching among mother-child dyads: Associations with child attachment and emotion reactivity. *Social Development* **26**, 610–629 (2017).

- [32] Rasmussen, H. F. *et al.* Mother-child language style matching predicts children’s and mothers’ emotion reactivity. *Behavioural Brain Research* **325**, 203–213 (2017).
- [33] Babcock, M. J., Ta, V. P. & Ickes, W. Latent semantic similarity and language style matching in initial dyadic interactions. *Journal of Language and Social Psychology* **33**, 78–88 (2014).
- [34] Müller-Frommeyer, L. C., Frommeyer, N. A. & Kauffeld, S. Introducing rls: An integrated metric assessing temporal reciprocity in language style matching. *Behavior Research Methods* **51**, 1343–1359 (2019).
- [35] Hove, M. J. & Risen, J. L. It’s all in the timing: Interpersonal synchrony increases affiliation. *Social cognition* **27**, 949–960 (2009).
- [36] Medio, A. & Lines, M. *Nonlinear dynamics: A primer* (Cambridge University Press, 2001).
- [37] Turvey, M. T. Coordination. *American psychologist* **45**, 938 (1990).
- [38] Webber Jr, C. L. & Zbilut, J. P. Recurrence quantification analysis of nonlinear dynamical systems. *Tutorials in contemporary nonlinear methods for the behavioral sciences* **94**, 26–94 (2005).
- [39] Wallot, S., Roepstorff, A. & Mønster, D. Multidimensional recurrence quantification analysis (mdrqa) for the analysis of multidimensional time-series: A software implementation in matlab and its application to group-level data in joint action. *Frontiers in psychology* **7**, 1835 (2016).
- [40] Coco, M. I. & Dale, R. Cross-recurrence quantification analysis of categorical and continuous time series: an r package. *Frontiers in psychology* **5**, 510 (2014).
- [41] Zbilut, J. P., Giuliani, A. & Webber Jr, C. L. Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Physics Letters A* **246**, 122–128 (1998).
- [42] Wallot, S. Multidimensional cross-recurrence quantification analysis (mdcrqa)—a method for quantifying correlation between multivariate time-series. *Multivariate behavioral research* **54**, 173–191 (2019).
- [43] Mankovich, A. *et al.* Say that again: Quantifying patterns of production for children with autism using recurrence analysis. *Frontiers in Psychology* **13**, 999396 (2022).
- [44] Brick, T. R., Gray, A. L. & Staples, A. D. Recurrence quantification for the analysis of coupled processes in aging. *The Journals of Gerontology: Series B* **73**, 134–147 (2018).

- [45] Duong, S., Davis, T. J., Bachman, H. J., Votruba-Drzal, E. & Libertus, M. E. Exploring dynamic structures of dyadic conversations using categorical cross recurrence quantification analysis: A tutorial. *The Quantitative Methods for Psychology* **20**, 121–136 (2024).
- [46] Likens, A. D., McCarthy, K. S., Allen, L. K. & McNamara, D. S. *Recurrence quantification analysis as a method for studying text comprehension dynamics*, 111–120 (2018).
- [47] Wallot, S. Recurrence quantification analysis of processes and products of discourse: A tutorial in r. *Discourse Processes* **54**, 382–405 (2017).
- [48] Leonardi, G. The study of language and conversation with recurrence analysis methods. *Psychology of Language and Communication* **16**, 165–183 (2012).
- [49] Jackson, E. S., Tiede, M., Riley, M. A. & Whalen, D. Recurrence quantification analysis of sentence-level speech kinematics. *Journal of Speech, Language, and Hearing Research* **59**, 1315–1326 (2016).
- [50] Boyd, R. L. in *Psychological text analysis in the digital humanities* 161–189 (Springer, 2017).
- [51] Sagan, C. A pale blue dot. <https://www.planetary.org/worlds/pale-blue-dot> (1994).
- [52] Pennebaker, J., Boyd, R., Booth, R., Ashokkumar, A. & Francis, M. Linguistic inquiry and word count: Liwc-22. pennebaker conglomerates. *Inc.* <https://www.liwc.app> (2022).
- [53] Darwin, C. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. *London: John Murray* (1859).
- [54] Fahnestock, J. *Rhetorical style: The uses of language in persuasion* (OUP USA, 2011).
- [55] Crossley, S. A. Linguistic features in writing quality and development: An overview. *Journal of Writing Research* **11**, 415–443 (2020).
- [56] Ta, V. P. *et al.* An inclusive, real-world investigation of persuasion in language and verbal behavior. *Journal of computational social science* **5**, 883–903 (2022).
- [57] Han, J. *Physiological and Emotional Synchrony During Television Viewing: A Nonlinear Dynamic Systems Approach* (Indiana University, 2020).
- [58] Tan, J.-W. *et al.* Repeatability of facial electromyography (emg) activity over corrugator supercilii and zygomaticus major on differentiating various emotions. *Journal of Ambient Intelligence and Humanized Computing* **3**, 3–10 (2012).

- [59] Okazaki, S., Yamanami, H., Nakagawa, F., Takuwa, N. & Kawabata Duncan, K. J. Mask wearing increases eye involvement during smiling: A facial emg study. *Scientific reports* **11**, 20370 (2021).
- [60] Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. lmerTest package: tests in linear mixed effects models. *Journal of statistical software* **82**, 1–26 (2017).
- [61] Brinberg, M. & Lydon-Staley, D. M. Conceptualizing and examining change in communication research. *Communication Methods and measures* **17**, 59–82 (2023).
- [62] Mogan, R., Fischer, R. & Bulbulia, J. A. To be in synchrony or not? a meta-analysis of synchrony’s effects on behavior, perception, cognition and affect. *Journal of Experimental Social Psychology* **72**, 13–20 (2017).
- [63] Tschacher, W., Rees, G. M. & Ramseyer, F. Nonverbal synchrony and affect in dyadic interactions. *Frontiers in psychology* **5**, 1323 (2014).
- [64] Abney, D. H., Paxton, A., Dale, R. & Kello, C. T. Complexity matching in dyadic conversation. *Journal of Experimental Psychology: General* **143**, 2304 (2014).
- [65] Dale, R., Duran, N. D. & Coco, M. Dynamic natural language processing with recurrence quantification analysis. *arXiv preprint arXiv:1803.07136* (2018).
- [66] Han, J., Cores-Sarria, L. & Zhou, H. In-person, video conference, or audio conference? examining individual and dyadic information processing as a function of communication system. *Journal of Communication* **74**, 117–129 (2024).
- [67] Solomon, D. H., Brinberg, M., Bodie, G., Jones, S. & Ram, N. A dynamic dyadic systems perspective on interpersonal conversation. *Communication Methods and Measures* **17**, 273–292 (2023).
- [68] Sherman, L. E., Michikyan, M. & Greenfield, P. M. The effects of text, audio, video, and in-person communication on bonding between friends. *Cyberpsychology: Journal of psychosocial research on cyberspace* **7** (2013).
- [69] Kitasuka, T., Aritsugi, M. & Rahutomo, F. *Semantic cosine similarity* (2012).
- [70] Gosling, S. D., Rentfrow, P. J. & Swann Jr, W. B. A very brief measure of the big-five personality domains. *Journal of Research in personality* **37**, 504–528 (2003).

Appendix A Numbered sentences from *On the origin of species*

(1) “When we look to the individuals of the same variety or sub-variety of our older cultivated plants and animals, one of the first points which strikes us, is, that they generally differ much more from each other, than do the individuals of any one species or variety in a state of nature.” (2) “When we reflect on the vast diversity of the plants and animals which have been cultivated, and which have varied during all ages under the most different climates and treatment, I think we are driven to conclude that this greater variability is simply due to our domestic productions having been raised under conditions of life not so uniform as, and somewhat different from, those to which the parent-species have been exposed under nature.” (3) “There is, also, I think, some probability in the view propounded by Andrew Knight, that this variability may be partly connected with excess of food.” (4) “It seems pretty clear that organic beings must be exposed during several generations to the new conditions of life to cause any appreciable amount of variation; and that when the organisation has once begun to vary, it generally continues to vary for many generations.” (5) “No case is on record of a variable being ceasing to be variable under cultivation.” (6) Our oldest cultivated plants, such as wheat, still often yield new varieties: our oldest domesticated animals are still capable of rapid improvement or modification.” (7) It has been disputed at what period of life the causes of variability, whatever they may be, generally act; whether during the early or late period of development of the embryo, or at the instant of conception.” (8) Geoffroy Hilaire’s experiments show that unnatural treatment of the embryo causes monstrosities; and monstrosities cannot be separated by any clear line of distinction from mere variations.” (9) “But I am strongly inclined to suspect that the most frequent cause of variability may be attributed to the male and female reproductive elements having been affected prior to the act of conception.” (10) “Several reasons make me believe in this; but the chief one is the remarkable effect which confinement or cultivation has on the functions of the reproductive system; this system appearing to be far more susceptible than any other part of the organisation, to the action of any change in the conditions of life.” (11) “Nothing is more easy than to tame an animal, and few things more difficult than to get it to breed freely under confinement, even in the many cases when the male and female unite.” (12) How many animals there are which will not breed, though living long under not very close confinement in their native country!” (13) “This is generally attributed to vitiated instincts; but how many cultivated plants display the utmost vigour, and yet rarely or never seed!” (14) “In some few such cases it has been found out that very trifling changes, such as a little more or less water at some particular period of growth, will determine whether or not the plant sets a seed.” (15) “I cannot here enter on the copious details which I have collected on this curious subject; but to show how singular the laws are which determine the reproduction of animals under confinement, I may just mention that carnivorous animals, even from the tropics, breed in this country pretty freely under confinement, with the exception of the plantigrades or bear family; whereas, carnivorous birds, with the rarest exceptions, hardly ever lay fertile eggs.” (16) “Many exotic plants have pollen utterly worthless, in the same exact condition as in the most

sterile hybrids.” (17) “When, on the one hand, we see domesticated animals and plants, though often weak and sickly, yet breeding quite freely under confinement; and when, on the other hand, we see individuals, though taken young from a state of nature, perfectly tamed, long-lived, and healthy (of which I could give numerous instances), yet having their reproductive system so seriously affected by unperceived causes as to fail in acting, we need not be surprised at this system, when it does act under confinement, acting not quite regularly, and producing offspring not perfectly like their parents or variable. “ (18) “Sterility has been said to be the bane of horticulture; but on this view we owe variability to the same cause which produces sterility; and variability is the source of all the choicest productions of the garden.” (19) “I may add, that as some organisms will breed most freely under the most unnatural conditions (for instance, the rabbit and ferret kept in hutches), showing that their reproductive system has not been thus affected; so will some animals and plants withstand domestication or cultivation, and vary very slightly—perhaps hardly more than in a state of nature.” (20) “A long list could easily be given of “sporting plants;” by this term gardeners mean a single bud or offset, which suddenly assumes a new and sometimes very different character from that of the rest of the plant.”