CrossMark

Introducing rLSM: An integrated metric assessing temporal reciprocity in language style matching

Lena C. Müller-Frommeyer[1] · Niels A. M. Frommeyer[1] · Simone Kauffeld[1]

Abstract

The way that individuals use function words in a conversation—reflecting how they say things, rather than what they say—is called their individual *language style*. The dyadic coordination of language styles, called *language style matching* (LSM), is central to the development of social relationships in conversations. Despite a growing body of research on LSM, conceptual and methodological approaches are inconsistent between scholars. After giving a conceptual overview of LSM, we derive the properties desirable for analyses of LSM in interaction (e.g., reciprocity, consistency, and frequency sensitivity). Building on these properties, the existing three methodological approaches to LSM are reviewed. Since none of the existing metrics fulfills all the desired properties, we introduce a new metric to assess LSM in dyadic interaction, capturing reciprocal adaption throughout the dynamic process of a conversation. Hence, the new metric is called *reciprocal LSM (rLSM)*. To empirically establish the conceptual underpinnings of rLSM, the metric is compared to the LSM metric most commonly used in psychological research. Both metrics are applied to a set of $N = 77$ transcribed real-life dyadic conversations, analyzed with the Linguistic Inquiry and Word Count software. The results indicate that rLSM is a better estimate of LSM than is the old metric and that there is high conceptual similarity between the two metrics. Implications for existing research and directions for future research are discussed. To facilitate the standardization and comparability of research, guidelines are provided for authors on the use of the new and existing metrics.

Keywords Language style matching · LIWC · Behavior coordination · Dyadic interaction

Language is an eminent part of human behavior and the basis of most social processes. By using language to communicate with one another, we share our emotions and thoughts and try to understand the world around us (Chung & Pennebaker, 2007). The most fundamental and widespread form of language use are verbal conversations. Conversations are defined as a progressive coordination of (linguistic) behaviors of at least two interlocutors (Pickering & Garrod, 2004). Conversations help us to build and foster social relationships by planning and coordinating with each other or by sharing our everyday experiences and memories (Clark, 1996; Tylén, Fusaroli, Bundgaard, & Østergaard, 2013). Because conversations are pervasively present, the question arises which part of the natural language within conversations influences our social relationships.

A young yet growing body of research on language style matching (LSM)—that is, the interdependent use of different types of function words (e.g., pronouns, articles, or prepositions)—shows that the coordination of function words influences our relationships with others (Ireland et al., 2011; Lord, Sheng, Imel, Baer, & Atkins, 2015; Niederhoffer & Pennebaker, 2002). Emphasizing function words rather than content words allows research to assess dyadic linguistic coordination irrespective of context. For example, two close friends, one of whom likes horse riding and the other of whom likes painting, would very likely use different content words when talking about their leisure time activities, whereas research has suggested that they would use function words in similar ways to the extent that they like and understand each other (this example was adapted from Ireland et al., 2011).

To date, LSM is used as a generic term for various conceptual and methodological approaches to the interpersonal coordination of language styles. To our knowledge, there are three different metrics to quantify LSM (Danescu-Niculescu-Mizil, Gamon, & Dumais, 2011; Ireland & Pennebaker, 2010; Niederhoffer & Pennebaker, 2002). There are also a number

✉ Lena C. Müller-Frommeyer
l.mueller-frommeyer@tu-bs.de

[1] Institute of Industrial/Organizational and Social Psychology, Technische Universität Braunschweig, Braunschweig, Germany

Springer

of other ways to quantify linguistic similarity, focusing on conversations' content rather than the language style. For example, Mikolov, Sutskever, Chen, Corrado, and Dean (2013) introduced Skip-gram analyses, a method to capture precise syntactic and semantic relationships in large bodies of natural language; Blei, Ng, and Jordan (2003) presented latent Dirichlet allocation, which uses Bayesian modeling to identify topical similarity in a body of natural-language samples; and Babcock, Ta, and Ickes (2013) used latent semantic similarity to analyze the contextual similarity between two natural-language samples. However, these methods are outside the scope of the present study, as we are interested in focusing on language style broadly rather than on content specifically.

Although all of the metrics used to analyze LSM have different advantages, they also have disadvantages, and the multitude of metrics makes it difficult to compare results across different studies. Furthermore, it is difficult for researchers to decide on the appropriate measure for their studies. Therefore, this article will provide a review of the conceptual and methodological approaches to the assessment of what is summarized under the term LSM. In doing so, we contribute to the literature by providing a comprehensive overview of the metrics used to assess LSM in dyadic interaction. By considering the conceptual underpinnings of LSM research, we suggest a set of properties that metrics assessing LSM should satisfy. Our subsequent review shows that none of the existing LSM metrics fulfills these properties. In line with Fusaroli, Raczaszek-Leonardi, and Tylén's (2014) call for metrics sensitive to the development and change of behavioral patterns over time, we see an additional contribution of our article in the introduction of an integrated metric to assess LSM in dyadic interaction. This metric complies with the desired properties and, thus, also considers the temporal reciprocity in the dynamics of conversations. We therefore call the new metric *reciprocal LSM* (rLSM). Finally, we statistically compare the new metric for LSM with the metric that is most prevalent in the psychological assessment of LSM, and derive guidelines for researchers to facilitate their decisions about the metric that is most appropriate for their research.

Unraveling LSM: What, why, and how?

Research has shown a long interest in the dyadic coordination of all kinds of behaviors—verbal and nonverbal alike—and their various outcomes (e.g., Chartrand & Bargh, 1999; Fusaroli et al., 2012). Since LSM is a relatively new phenomenon under investigation, we will subsequently describe the basics of LSM research. First, we clarify the terminology used in this article to describe the phenomenon and provide explanations of *what* language styles are. Then we describe the theoretical underpinnings, explain *why* LSM is a relevant phenomenon in the field of behavior coordination, and provide

examples of *how* language styles are quantified, in order to further analyze the coordination thereof.

Clarifying the what—Terminology and language styles

As Paxton and Dale (2013) acknowledged, the terminology in the field of interpersonal coordination is scattered. To facilitate the reading of this article, we will shortly name and explain the terminology that we will apply. With LSM being central to this article, we focus on the dyadic coordination of function words over the process of a conversation (Niederhoffer & Pennebaker, 2002). By using terms such as *coordination* and *accommodation*, we imply that the process—that is, the reciprocal development and change of linguistic behaviors—is under consideration (Fusaroli et al., 2014). In contrast, we will use the term *similarity* when the focus moves away from process and texts are instead treated as static (Ireland et al., 2011; Ireland & Henderson, 2014).

To fully understand the concept of LSM, it is essential to know what the *language style* in language style matching stands for. Each person has a unique language style, represented by the specific individual use of function words (e.g., pronouns, articles, and prepositions) (Pennebaker, Mehl, & Niederhoffer, 2003), describing *how* people say things rather than *what* they say. Even though function words account for 60% of the words we utter (Pennebaker, 2013), they are short and have little meaning outside the context of a sentence (Chung & Pennebaker, 2007). As such, their use is more automated and nonconscious than the use of content words such as nouns and verbs (Segalowitz & Lane, 2004). Furthermore, function words alone do not reveal any specific content. Therefore, a shared social knowledge and understanding of the conversational topic is needed among the conversational partners to correctly understand and respond to them. For example, the function words in the sentence "They played it successfully" (*they* and *it*) only make sense for conversational partners who have prior knowledge of the group (*they*) and the object (*it*) in question. If two people engage in a conversation and share a common ground—that is, common goals or information—they are more likely to coordinate their (especially linguistic) behaviors (Brennan & Hanna, 2009; Paxton, Dale, & Richardson, 2016). Therefore, the successful coordination of function words reflects a common understanding of the conversational topic and a shared social knowledge (Meyer & Bock, 1999).

Attending to the why—Theoretical underpinnings of LSM

The theoretical foundations of LSM are manifold—ranging from interpersonal coordination (Bernieri & Rosenthal, 1991; Lumsden, Miles, Richardson, Smith, & Macrae, 2012)

and behavioral mimicry (Lakin & Chartrand, 2003), in the domain of motor coordination, to communication accommodation theory (Giles & Coupland, 1991; Shephard, Giles, & LePoire, 2001) and interactive alignment (Muir, Joinson, Cotterill, & Dewdney, 2016; Pickering & Garrod, 2004), in the field of linguistic coordination. These foundations share conceptual basics, especially the belief that the coordination of behaviors is automatic and leads to mostly positive outcomes, such as increased rapport (Chartrand & Bargh, 1999). Additionally, these conceptual similarities make it hard to distinguish the different theoretical approaches (Paxton & Dale, 2013).

The idea of automatic coordination is challenged by the relatively new interpersonal synergy approach (Riley, Richardson, Shockley, & Ramenzoni, 2011), which emphasizes the coordination of different linguistic systems in order to commonly fulfill tasks at hand (Paxton et al., 2016). Just as specific muscles in a body need to coordinate in order to achieve a common goal (e.g., to move a foot), specific features of two interlocutors' linguistic systems coordinate in order to solve mutual tasks (e.g., when discussing a problem; Paxton et al., 2016). Growing theoretical (Fusaroli et al., 2014; Paxton et al., 2016) and experimental (Fusaroli et al., 2012; Tylén et al., 2013) evidence in the motor as well as the linguistic domain has shown that coordination is context-sensitive, adaptive, and not always linked to positive outcomes. Fusaroli et al. (2012) showed that, to solve a dyadic perceptual task, a general accommodation of linguistic features did not have a positive effect on performance, whereas the accommodation of specific task-relevant vocabulary was positively linked to performance.

Despite the disagreement on the question of automatic coordination, scholars agree that coordination is mostly nonconscious in the motor as well as the linguistic domain (Fusaroli et al., 2014). Although most studies on LSM support the idea that coordination yields positive outcomes—for example, relationship initiation and stability (Ireland et al., 2011), empathy between clients and therapists (Lord et al., 2015), and peaceful resolutions of hostage negotiations (Taylor & Thomas, 2008)—other studies support the interpersonal synergy approach, by providing evidence that LSM is context-sensitive (Babcock et al., 2013; Bowen, Winczewski, & Collins, 2017) and does not necessarily reflect interpersonal rapport, but rather intensifies the positive or negative atmosphere of an interaction. Using the evidence at hand, we build on the interpersonal synergy approach and conceptualize the coordination of language styles (i.e., LSM) as a nonconscious linguistic behavior that occurs when interacting with at least one interactional partner in order to achieve any kind of common goal (context-sensitive).

Building on our conceptualization of LSM and the definition of conversations as the progressive coordination of linguistic behaviors, methods used to represent the process of coordination need to be able to detect and display the change

and development of coordination over time (Fusaroli et al., 2014). Scholars agree that behavioral coordination unfolds over time and is reciprocal in nature (e.g., Abney, Paxton, Dale, & Kello, 2014; Hove & Risen, 2009; Main, Paxton, & Dale, 2016). Hence, an accommodation of language style can only be investigated accurately if its process and the interlocutors' reactions to the preceding statements are considered at any given time (e.g., Fusaroli et al., 2014).

Summarizing, LSM and its effects on social interaction are classified as being relevant for many parts of human social interactions involving natural language. Because LSM is essentially reciprocal, analyses cannot be conducted independently of either the interactional partners or the processes over time that underlie the conversations.

Explaining the how—Analyzing language styles

The basis of LSM research is the natural language used by each of the conversational partners. All words uttered within a conversation need to be quantified in order to make them accessible for statistical analyses. To gain a better understanding of the process of language style analysis, we will subsequently illustrate it with the help of an example taken from William Shakespeare's (1597) tragedy *Romeo and Juliet*.

To start with, imagine yourself sitting on your sofa in front of your television or in a theater, reading or watching *Romeo and Juliet* (Shakespeare, 1597). When you get to the short scene in Table 1, in which Romeo talks to his cousin Benvolio, you suddenly feel the urge to assess the language style of both characters and start wondering *how?*.

Preparing the analyses In a first step, the scene is transcribed word for word, clearly marking each speaker (see Table 1 for the transcript). After the transcription are completed, the text file needs to be quantified using appropriate software. All LSM research published so far has been based on the software Linguistic Inquiry and Word Count (LIWC; Pennebaker, Booth, & Francis, 2007; Pennebaker, Boyd, Jordan, & Blackburn, 2015), which is why we will subsequently explain the process of language style analyses using LIWC. With the increasing visibility of open-source projects, it is important to note that LIWC is neither a free nor an open-source tool. The software reads all the words in a given text file and compares them with a built-in dictionary. For example, in its current version, the complete English[1] dictionary contains 6,400 entries, which are all assigned to one or more of 55 nonexclusive categories, organized under four main themes: basic linguistic processes, psychological processes, personal concerns, and spoken categories (Pennebaker et al., 2015).

---

[1] Dictionaries for other languages (e.g., German, French, and Italian) can be downloaded from the LIWC website (www.liwc.net/dictionaries).

Table 1   Transcript of a short dialogue from Shakespeare's tragedy *Romeo and Juliet* (Shakespeare, 1597)

Speaker	Statement
Romeo	*Is the* day *so* young?
Benvolio	*But* new struck nine.
Romeo	Ay *me*! Sad hours seem long. *Was that my* father *that* went *hence so* fast?
Benvolio	*It was. What* sadness lengthens Romeo's hours?
Romeo	*Not having that*, *which*, *having*, makes *them* short.

Function words are written in *italics*

The analysis can be run for all categories across the whole transcript or for a user-defined selection of categories across certain levels of analysis within a transcript, ranging from whole text to single-word analyses. Depending on your personal research question, it is advisable to choose an individual level of analysis that suits your needs. For all words in a chosen level, the software reports the proportions of the chosen categories. The analysis of language style only focuses on the so-called function word categories, containing pronouns, articles, prepositions, auxiliary verbs, adverbs, conjunctions, and negations in the English dictionary (Pennebaker et al., 2015).

Different levels of language style analysis To illustrate the possibilities among different levels of analysis, we subsequently report results of the language style analyses on different levels of analysis with the help of the *Romeo and Juliet* example. The results can be found in Table 2.

In Example A, the researcher is interested in the language style across the whole excerpt of the conversation, not distinguishing between speakers. The results indicate that across the whole transcript, the speakers used 39 words altogether, of which 89.74% were recognized by the software. From the total words, 51.28% were function words, which can be further assigned to the individual function word categories—for example, 23.08% fall in the pronoun category, 2.56% fall in the article category, and 0% fall in the prepositions category.

In Example B, the researcher might be interested in each speaker's individual language style across the whole document. To provide results for Example B, the transcript was separated by speaker prior to the analyses, producing one document per speaker. The results show that Speaker A (Romeo) used 28 words across the excerpt of the conversation, of which 96.43% were recognized by the software. Of the total words, 57.14% fall in the function word category. Speaker B (Benvolio), on the other hand, used 11 words across the excerpt of the conversation, of which 72.73% were recognized by the software. Of the total words, 36.36% fall in the function word category. This level of analysis allows for comparisons between speakers and yields that Romeo talked a lot more during the excerpt of the conversation than did his cousin

Benvolio. Additionally, Romeo used nearly twice as many function words as Benvolio.

In Example C, each statement of the conversation was analyzed with the LIWC. In this case, the researcher might want to learn about each speaker's individual language style in each statement across the whole conversation. The results show that in the first statement, Speaker A used five words that were all recognized by the software, and 60% fall in the function word category. In his first statement, Speaker B uses four words, of which 75% were recognized by the software and 25% fall in the function word category. Taking a closer look at each statement, one can see the variance in the use of function words. If you compare the analyses made in Examples B and C, you can see that a more fine-grained analysis of the different function word categories reveals more fluctuations in language use between the two characters.

Table 2   LIWC results for different units of analysis in the "Romeo and Juliet" example

Level of analysis	WC	Dic (%)	FW (%)	Pron (%)	Art (%)	Preps (%)
Example A						
Whole text	39	89.74	51.28	23.08	2.56	0
Example B						
Speaker A (Romeo)	28	96.43	57.14	25	3.57	0
Speaker B (Benvolio)	11	72.73	36.36	18.18	0	0
Example C						
Speaker A – Statement 1	5	100	60	0	20	0
Speaker B – Statement 1	4	75	25	0	0	0
Speaker A – Statement 2	15	93.33	46.67	26.67	0	0
Speaker B – Statement 2	7	71.43	42.86	28.57	0	0
Speaker A – Statement 3	8	100	75	37.5	0	0

A selection of language style categories, to illustrate the process of language style analyses. WC = word count (total number of words); Dic = dictionary (proportion of words recognized by the LIWC); FW = function words (proportion of function words in the level under investigation); Pron = pronouns (proportion of pronouns in the level under investigation); Art = articles (proportion of articles in the level under investigation); Preps = prepositions (proportion of prepositions in the level under investigation)

Choosing the appropriate level of analysis Taken together, the chosen levels of analysis reveal different aspects of the conversation at hand. Thus, depending on the focus of your research, it is possible to receive summary scores as well as very fine-grained scores for each conversation. For example, if you were interested in comparing the frequencies of function word usage in the Romeo–Benvolio dyad with another dyad, you might consider the whole conversation (Example A). To investigate Romeo's stability of function word use across conversations with different partners, you would more likely use a speaker-separated analysis (Example B). Finally, to assess the coordination of function words between Romeo and Benvolio in the present conversation, you would need a fine-grained analysis that reveals the time course of the conversation, opting for the talk-turn-based analysis of specific function word categories (Example C).

Desirable properties for metrics assessing LSM

After assessing Romeo's and Benvolio's language styles, the next step would be to assess both characters' matching of language styles. Before presenting the existing metrics used to assess LSM, we will describe desirable properties for LSM metrics, derived from the conceptual and methodological underpinnings presented so far. Table 3 provides a summary of these properties.

Property 1: Stylistic analyses

First, in line with our previous explanations, language *style* matching focuses on the coordination of language styles. As such, only those LIWC categories that fall into the theme of language style should be included in the LSM analysis. These are summarized under the LIWC 2015 function word scores (i.e., pronouns, articles, prepositions, auxiliary verbs, adverbs, conjunctions, and negations). Even though the exclusive use of language style categories should be self-evident when

analyzing LSM, previous scholars have included nonstylistic categories (e.g., causation or insight) into their calculation of LSM (Danescu-Niculescu-Mizil et al., 2011; Taylor & Thomas, 2008). Acknowledging the fact that not all of the function word categories are available in the dictionaries of all languages, an adaption according to the particular dictionary is acceptable.

Property 2: Reciprocity

Second, to account for the reciprocity of conversations and behavioral accommodation, metrics for LSM in natural conversations need to consider the temporal dynamics of conversations. Thus, the words Speaker A uses in Statement 1 influence the words Speaker B uses in Statement 1, whereas Speaker B's choice of words in Statement 1 influences Speaker A's choice of words in Statement 2, and so forth.

Property 3: Flexibility

Third, to allow for the analysis of individual research questions (e.g., language style differences between speakers vs. language style differences between conversations), the metric needs to provide flexibility as to each of its three constituent parts: (1) the two speakers who interact as a dyad, (2) the statements of a conversation, and (3) the LIWC categories previously defined as stylistic. Therefore, the metric should allow for the calculation of results for each individual speaker to uncover, for example, leader–follower dynamics in the process of LSM, as well as one dyadic score to allow for between-dyad comparisons. Furthermore, the metric needs to be adaptable in its consideration of the temporal dynamics of conversations; that is, it should provide the possibility to include a predefined set of consecutive statements. By doing so, the metric can allow us to investigate LSM over the course of a whole conversation, individually identified phases, or only a short exchange of words. For the third component—the language style categories—these should provide one average

Table 3 Summary of desirable properties for LSM metrics

Desired Property	Description
(1) Stylistic Analyses	The metric exclusively includes language style categories.
(2) Reciprocity	The metric includes the temporal dynamics of the conversation.
(3) Flexibility	The selection of speakers, temporal dynamics, and language style categories included in the LSM score can be made depending on the personal research question.
(4) Consistency	The properties of the LSM score is consistent and independent of the parameters chosen for Property 3 (flexible)—that is, the score always ranges from 0 to 1, allowing for meaningful interpretations.
(5) Frequency sensitivity	The metric considers differences between low- and high-frequency categories.
(6) Replicability	The calculation of the metric is replicable by, and thereby accessible to, researchers in the field.

LSM score and, if theoretically needed, one score for each of the language style categories.

Property 4: Consistency

Fourth, despite the abovementioned flexibility in the parameters of the analysis, the metric itself (i.e., the equation utilized) should be applicable consistently across all types of analyses, from short statements to whole conversations. Therefore, supporting Guastello and Peressini's (2017) call for a "measure of the size of synchronization effect, not simply a significance test" (p. 9), the scores derived should quantify the actual behavior into a classifiable score, which will allow for meaningful interpretations and comparisons.

Property 5: Frequency sensitivity

Fifth, metrics for LSM need to consider frequency differences in the use of categories. As Example C in Table 2 illustrates, the use of function word categories may fluctuate across statements. If, for example, prepositions make up 4% of a conversation, a difference of two percentage points between two speakers will be more meaningful than a two-percentage-point difference in a category that makes up 20% of the conversation (Chung & Pennebaker, 2007). Hence, LSM metrics should consider the relative use of specific function words in the respective level of analysis (Ireland & Pennebaker, 2010).

Property 6: Replicability

Sixth, to standardize research on LSM and reduce the degree of complexity, metrics should provide guidelines for replication, and thereby be accessible to other researchers.

Review of the existing metrics to assess LSM in dyadic interaction

In the following section, we present the existing metrics for the assessment of LSM and evaluate them with regards to the desirable properties described above. Since all of the studies considered propose different metrics for the same construct, we will further use the authors' names to distinguish between the metrics.

Niederhoffer and Pennebaker (2002)

The first metric to assess LSM in two documents was introduced by Niederhoffer and Pennebaker (2002). Their aim was to uncover dyadic dynamics, such as leader–follower dynamics in LSM, by investigating the degree to which conversational partners match their language styles. To do so, they introduced two different forms of analyses. First, they

provided a between-subjects analysis, in which they correlated the degree to which Speaker A uses a specific choice of words (e.g., prepositions) throughout a whole conversation with the degree to which Speaker B uses the same choice of words. The calculation of the LSM metric can therefore be represented by the following Eq. (1):

$$LSM_{AB}(C) = Cor(C_A, C_B), \tag{1}$$

where C is the respective language style category and A and B are both speakers in a dyad.

As was already stated, the proposed equation only allows for a between-dyad comparisons. To further uncover leader–follower patterns of LSM, Niederhoffer and Pennebaker (2002) expanded their metric. To account for the reciprocal influence of the speakers' statements—Statement 1 by Speaker A impacts Statement 1 by Speaker B, which in turn impacts Statement 2 by Speaker A, and so forth—another two sets of correlations are computed within each conversation.

In the first set, LSM is calculated for each talk-turn and category of Speaker B by correlating all statements of Speaker B following statements of Speaker A (Eq. 2; for example, the proportion of prepositions used by Speaker B in Statement 1 is correlated with the proportion of prepositions used by Speaker A in Statement 1). Similarly, in the second set, LSM is calculated for each talk-turn and category of Speaker A. Therefore, the statements of Speaker A are lagged by one (Eq. 3; for example, the proportion of prepositions used by Speaker B in Statement 1 is correlated with the proportion of prepositions used by Speaker A in Statement 2). Final LSM scores are obtained by calculating the mean of Speaker A's and B's individual LSM (Eq. 4)—that is, how Speaker A matches his/her language style to Speaker B, and vice versa. The scores can take values between – 1 and 1 and are interpreted like correlations.

$$LSM_B^{S=1}(C) = Cor(C_A^{S=1}, C_B^{S=1}), \tag{2}$$

$$LSM_A^{S=1}(C) = Cor(C_B^{S=1}, C_A^{S=2}), \tag{3}$$

$$LSM_{AB}^S(C) = M(LSM_A^S, LSM_B^S), \tag{4}$$

where A and B are both speakers in the dyad, S is the statement considered in the analyses, and C is the respective function word category.

Compliance with the desired properties In summary, the metric presented considers Property 1 by including only stylistic categories. Property 2 is also fulfilled by considering the reciprocity of dyadic interactions. Furthermore, it provides opportunities for flexible but consistent calculations of LSM scores (Properties 3 and 4), which have been adapted to the analyses of LSM in different phases in hostage negotiations (Taylor & Thomas, 2008). However, since some methodological considerations speak against the application of

correlations to assess LSM, this metric does not completely comply with Property 4. First, for assessing LSM in the way described above, the first step is to compute numerous correlations—one for each speaker, talk-turn, and category—and using these correlations in the following steps. This approach applies as long as a correlation is significant. The correct procedure would exclude nonsignificant (presumably low) results from the following analysis, potentially resulting in LSM scores that are artificially high in magnitude (Niederhoffer & Pennebaker, 2002). Additionally, the metric is not frequency-sensitive (Property 5), because the correlations are based on the absolute usage of single categories, disregarding their relative occurrence in the unit of analyses. Because the authors thoroughly describe the procedure applied to calculate the LSM score and use basic statistical methods to do so, it can be considered accessible, thereby fulfilling Property 6.

Ireland and Pennebaker (2010)

Ireland and Pennebaker (2010) introduced the metric most frequently used in psychological research on LSM in dyadic conversations (e.g., Cannava & Bodie, 2016; Ireland et al., 2011; Rains, 2015). In this metric, a weighted difference score is calculated for each LSM category. First, in the numerator, the absolute value of the difference between the LIWC results within a specific language style category (C) for Speaker A (C_A) and Speaker B (C_B) is calculated, which is then—in the denominator—divided by the sum of C_A and C_B. In the denominator, .0001 is added in order to prevent the empty sets that would occur if the value for the category in question were 0% at each level of analysis. Then, the result of this fraction is subtracted from 1 (Eq. 5), resulting in a value between 0 and 1, with higher values indicating higher LSM in the respective categories.

$$LSM_{AB}(C) = 1 - \frac{|C_A - C_B|}{C_A + C_B + 0.0001} \qquad (5)$$

Finally, LSM scores for each of the function word categories are averaged to yield a composite LSM score. Like the category scores, LSM can take a score between 0 and 1, with higher scores representing higher LSM. Note that when neither of the speakers uses any given LSM category, then the LSM score for that specific category will be 1. Therefore, language style categories that are not used in a chosen conversation are excluded from the LSM calculation. By considering each category's overall frequency, it is ensured that differences in less-frequent categories have a greater impact on the overall LSM score than the same absolute difference in higher-frequency categories (Ireland & Pennebaker, 2010).

Compliance with the desired properties Even though this metric was originally introduced in order to calculate LSM in nondynamic written texts, it is also commonly used for assessing LSM in dyadic conversations. The conceptual analysis of this metric shows that it is frequency-sensitive (Property 5) and accessible to other researchers (Property 6). However, when applied to dyadic interaction, this metric does not consider the reciprocity of conversations (Property 2). Neither does it allow for flexibility in the calculation of individual LSM scores, the particular level of analysis, nor does the metric allow for a consistent quantification of LSM across all possible levels of analysis (Properties 3 and 4).

Danescu-Niculescu-Mizil et al. (2011)

A third metric to calculate linguistic accommodation, which shows conceptual similarities to LSM, was introduced by Danescu-Niculescu-Mizil et al. (2011). The metric was developed to analyze linguistic accommodation in big data, such as online conversations between two conversational partners on Twitter. The authors analyzed a data set of more than three million talk turns between 2,200 pairs of users. Prior to their analyses, the online conversations were quantified using the LIWC software. Linguistic style accommodation is evaluated using a probabilistic framework that specifically accounts for the temporal dynamics of the conversations. The probabilistic framework looks as follows:

$$Acc_{(A,B)}(C) \triangleq P\left(T_B^C \mid T_A^C, T_B \hookrightarrow T_A\right) - P\left(T_B^C \mid T_B \hookrightarrow T_A,\right) \qquad (6)$$

where A and B represent a pair of users who hold a conversation on Twitter, C represents any stylistic category of the LIWC, and T is short for tweet/statement. T_A^C and T_B^C represent tweets containing a certain stylistic category posted by one of the users. The arrow ($\hookrightarrow$) represents the answer condition—that is, in $T_B \hookrightarrow T_A$, User B replies to User A. Hence, the minuend describes the probability that User B's reply to User A contains a certain stylistic category under the condition of it being an answer to the previous tweet of A. The subtrahend denotes the probability that User B's tweet to A would contain C under normal circumstances—that is, without it being a reaction to a previous tweet. Thus, it assesses how much the event of User A using C in a tweet affects the probability of User B using C in a reply to User A, above the normally expected use of C by User B. Finally, the following Eq. 7 captures the accommodation for a given category C throughout the entire online conversation:

$$Acc(C) = E\left[Acc_{(A,B)}(C)\right], \qquad (7)$$

where the expectation E is calculated over the language style accommodation of Users A and B across all possible

consecutive pairs of tweets in an online conversation. Given this framework, $Acc(C) > 0$ reflects language style accommodation in C.

Compliance with the desired properties By modeling language style accommodation over the course of each statement of an online conversation, this approach considers the users' reciprocity (Property 2) and provides full flexibility in LSM calculation (Property 4). Even though the approach allows for consistent quantification of LSM and provides a score for each user, it is not frequency-sensitive, because the probabilities only focus on the occurrence of categories, regardless of their relative occurrence. Thus, the higher impact of differences in lower-frequency categories, in comparison to differences in higher-frequency categories, is not considered (Property 5). Finally, the replicability of this metric is limited—most probably due to the sophisticated probabilistic approach—which is represented by the single psychological study that has used this metric to assess LSM (Lord et al., 2015).

Concluding summary of existing LSM metrics

In summary, the three approaches to assess LSM in dyadic interaction each fulfill some but not all of the desired properties. An overview of the metrics and their fulfillment of the properties can be found in Table 4. Nonetheless, each of the metrics presented is appropriate for its originally designated context. Since none of the presented metrics considers all of the desired properties, we will subsequently present an integrated metric to assess LSM that builds on the metrics presented so far.

Introducing an integrated metric to assess LSM in dyadic interaction–Reciprocal LSM (rLSM)

Our new metric to assess LSM in dyadic interaction is based on and extends the metric by Ireland and Pennebaker (2010) and, moreover, considers the dynamic, talk-turn-based nature of conversations, as proposed by Niederhoffer and Pennebaker (2002) and Danescu-Niculescu-Mizil et al. (2011). After presenting guidelines for the treatment of missing values, we will explain the process of calculating rLSM for any one of the respective stylistic LIWC categories (C) using the new metric, and its compliance with the desired properties with the help of the *Romeo and Juliet* (Shakespeare, 1597) example. We provide an R script for the calculation of the rLSM metric as supplemental material of this article.

Treatment of missing values

When one is working with the statement-based LIWC output, missing values need to be considered and treated in a specific way. Missing values can occur whenever a given language style category is not used in at least one conversational partner's statement. In the LIWC output, these cases are denoted with a zero. Three different combinations of missing values in subsequent statements are possible (see Table 5).

In Example A, Speaker A does not use any words that fall in the LIWC category of prepositions in his statement, and neither does Speaker B. Not replacing the zeros with missing values would result in an rLSM score of 1—which would

Table 4 Property overview for each of the presented metrics and respective recommendations of use

Source	Niederhoffer & Pennebaker (2002)	Ireland & Pennebaker (2010)	Danescu-Niculecu-Mizil et al. (2011)	Müller-Frommeyer, Frommeyer, & Kauffeld (This study)
Property 1: *Stylistic analyses*	●	◑	◑	●
Property 2: *Reciprocity*	●	○	●	●
Property 3: *Flexibility*	●	○	●	●
Property 4: *Consistency*	◑	○	●	●
Property 5: *Frequency sensitivity*	○	●	○	●
Property 6: *Replicability*	●	●	◑	●
Recommendations for use	Use not recommended, as advantages of other metrics outweigh	Static, independent texts (e.g., comparison of books, poetry or speeches)	Written online conversations (e.g., Twitter)	Reciprocal, dynamic verbal processes (e.g., dyadic or group conversations)

● Property fulfilled. ◑ Property partly fulfilled. ○ Property not fulfilled

Table 5 Examples for the replacement of zeros in the LIWC data output

	Data Before Replacement		Data After Replacement	
	Speaker A: Preps	Speaker B: Preps	Speaker A: Preps	Speaker B: Preps
Example A	0	0	NA	NA
Example B	0	5	NA	5
Example C	5	0	5	0

Preps = prepositions. NA = missing value

reflect perfect synchrony but not a nonobservation of behaviors. In Example B, Speaker A does not use prepositions in his statement, but Speaker B does. Since there is no behavior that Speaker B could follow, Speaker A's zero needs to be replaced by a missing value, so that no LSM score is calculated. In Example C, however, Speaker A does use prepositions, whereas Speaker B does not, even though B had the opportunity to follow A's behavior. Here, the zero is not replaced, and instead the resulting rLSM score for *prepositions* in this statement is low.

Disregarding these assumptions would lead to both upward and downward biases in rLSM scores. Combinations like Example A would lead to very high matching scores, and combinations like Example B would lead to very low scores, whereas both should not affect the matching scores at all.

Calculation of statement-based rLSM scores

To begin with, we illustrate the calculation of rLSM with the help of the smallest reasonable unit of analysis: two consecutive statements. Referring to the *Romeo and Juliet* example, we will use Statement 1 by Romeo—subsequently referred to as Speaker A—and Statement 1 by Benvolio—subsequently referred to as Speaker B. Because this consecutive pair of statements captures Speaker B's reaction to Speaker A, Speaker B's rLSM is represented by the following equation:

$$rLSM_B(C) = 1 - \frac{\left|C_A^{S=1} - C_B^{S=1}\right|}{C_A^{S=1} + C_B^{S=1} + 0.0001}, \tag{8}$$

where S indicates the respective statement—that is, the level of analysis. Note that C can be any LIWC category defined as stylistic. If applied to the *Romeo and Juliet* example, with C representing the LIWC results of the function word (FW) category, we can see that $rLSM_B(FW) = .59$. To fully represent the reciprocity of the conversation and calculate Speaker A's rLSM to Speaker B, Speaker A's second statement in response to Speaker B needs to be considered (Eq. 9).

$$rLSM_A(C) = 1 - \frac{\left|C_B^{S=1} - C_A^{S=2}\right|}{C_B^{S=1} + C_A^{S=2} + 0.0001} \tag{9}$$

Applied to the *Romeo and Juliet* example, with C being the LIWC results of the function word category, Speaker A gets a score of $rLSM_A(FW) = .70$. The results indicate that Speaker A (Romeo) exhibits more LSM than does Speaker B (Benvolio).

To be able to capture the temporal dynamics of rLSM across the whole conversation, the equation needs to be expanded in its levels of analysis—that is, the temporal dimension S needs to be generalized to be applicable to any statement in the conversation (i). This expansion results in the following two equations:

$$rLSM_A(C) = 1 - \frac{\left|C_B^{S=i} - C_A^{S=i+1}\right|}{C_B^{S=i} + C_A^{S=i+1} + 0.0001}, \tag{10}$$

$$rLSM_B(C) = 1 - \frac{\left|C_A^{S=i} - C_B^{S=i}\right|}{C_A^{S=i} + C_B^{S=i} + 0.0001}. \tag{11}$$

Calculation of conversation-based individual rLSM scores

To retrieve individual rLSM scores for the whole conversation, the equation needs to be further expanded in its temporal dimension S. That is, in addition to the application to one statement, the metric needs to be applicable to a predefined set of statements ($S = i - k$). To produce one rLSM score per speaker for the defined set of statements, the LSM scores per speaker are averaged across statements (Eqs. 12 and 13). With this expansion, it is possible to calculate individual rLSM scores per category for predefined phases of a conversation and for whole conversations.

$$rLSM_A^{S=i-k}(C) = M\left(rLSM_A^{S=i-k}(C)\right) \tag{12}$$

$$rLSM_B^{S=i-k}(C) = M\left(rLSM_B^{S=i-k}(C)\right) \tag{13}$$

In the *Romeo and Juliet* example, across all statements, Romeo shows an overall rLSM score of $rLSM_A^{S=1-3}(FW) = .77$, whereas Benvolio shows an overall rLSM score of $rLSM_b^{S=1-3}(FW) = .71$, still indicating that Romeo exhibits more LSM than Benvolio.

Calculation of dyadic rLSM scores

The calculation of one final dyadic rLSM score is reached by the integration of the abovementioned equations (10–13), which is represented by Eq. 14. The dyadic rLSM score considers both speakers (A, B) and the defined set of statements $(S = i - k)$—that is, all consecutive pairs of statements in a conversation.

$$rLSM_{AB}^{S=i-k}(C) = M\left(rLSM_{A}^{S=i-k}(C), rLSM_{B}^{S=i-k}(C)\right) \quad (14)$$

Because this equation is still based on only one LIWC language style category at a time, the results for all of the language style categories need to be averaged in order to calculate the final rLSM score. For the *Romeo and Juliet* example, the final LSM score is $rLSM_{AB}^{S=1-3}(FW) = .74$. By considering all of the available function word categories, the new metric can be classified as exclusively stylistic (Property 1). Furthermore, by calculating individual scores per speaker and pairs of statements, the metric further considers the reciprocity and temporal dynamics of the conversation (Property 2). Because categories, speakers, and the amount of consecutive statements can be defined individually, the metric also provides the desired flexibility (Property 3) but is still consistent in its calculations and, therefore, allows for meaningful comparisons between scholars (Property 4). Since the metric is based on Ireland and Pennebaker's (2010) frequency-weighted difference score, it can also be classified as frequency-sensitive (Property 5). The stepwise calculation of rLSM is based on the LIWC output with the possibility for individual adaption of the three constituent parts, and thus it can be replicated by using the R script in the additional online material and the integrated instructions. Therefore, the metric also fulfills Property 6.

Empirical examination of the rLSM metric

In addition to the conceptually and methodologically based introduction of the rLSM metric, we will statistically investigate the underpinnings presented. Therefore, we compare the rLSM metric to the LSM metric introduced by Ireland and Pennebaker (2010), which is the most dominant metric in psychological research (e.g., Ireland & Henderson, 2014; Ireland et al., 2011; Rains, 2015).

The rLSM metric considers the mutual use of language style words on a talk-turn level, whereas the LSM metric assesses the similar use of language style on the broader, conversational level. This resembles other methodological areas, such as interrater reliability calculations, in which point-by-point agreements provide lower, more accurate, and therefore more realistic estimates of agreement than do simple overall agreements (Klonek, Quera, & Kauffeld, 2015). Since the

likelihood of finding corresponding language style types is inflated in the LSM metric, we expect rLSM scores to be lower than LSM scores (Niederhoffer & Pennebaker, 2002; Hypothesis 1).

Both rLSM and LSM are based on the same conceptual underpinnings and are supposed to fulfill the same social functions. Furthermore, since the LSM metric captures similarity in the use of function words on the conversational level, ignoring its temporal dynamics—whereas rLSM captures the accommodation of function words unfolding over adjacent statements of a conversation—and since both are related in their fundamentals, we argue that they cover similar concepts. Therefore, we expected to find a positive relation between the metrics for rLSM and LSM (Hypothesis 2). We expected this relationship to be present for each language style category as well as for the overall rLSM and LSM scores.

Method

Participants

The analyses reported here include interactions of 77 same- and mixed-sex dyads (male = 112, female = 41; one person did not indicate their sex). The mean age was 23.81 years (SD = 3.28). All participants voluntarily participated in an exercise to get to know each other by holding short conversations with a previously unacquainted partner. The majority of the participants studied engineering ($n = 127$), whereas the remaining ones ($n = 27$) indicated different academic fields of study, including psychology, history, biology, chemistry, or sociology. Altogether, the sample represents 145,123 words ($M = 1,884.71$, $SD = 921.70$) and more than 12,500 talk turns, which exceeds the data usually used in LSM research (e.g., Ireland et al., 2011; Niederhoffer & Pennebaker, 2002).

Procedures

In total, 164 participants (82 dyads) took part in a study designed to investigate the relationship between LSM and relationship initiation in formerly unacquainted dyads over the course of several conversations. The study was integrated into interdisciplinary soft-skill trainings on communication and conflict management at a German university. Participants signed up for the trainings via the university homepage. Participation in the trainings was voluntary and compensated with class credit. At the beginning of the training, participants were randomly assigned to a conversational partner they had not previously been acquainted with. They conducted two different conversations with their partner: One on basic personal aspects, such as family, education, or hobbies, and the other one on a conflicting topic. Because LSM has been shown to be context-sensitive (Bowen et al., 2017), we only included the conversations on basic personal aspects in our

analyses. Participants voluntarily audiotaped their conversations for scientific purposes. Five of the dyads did not audiotape their conversations, resulting in a final sample of 154 participants (77 dyads). After the conversations, participants presented the most important facts about their respective partner to the rest of the group, to enhance group cohesion before the training continued. Participants were not compensated separately for participating in the study. All participants provided written consent to be audiotaped, and all procedures of the study were approved by the institutional review boards on data security and ethics.

Measures

Language style matching

To compare rLSM and LSM scores in real-life conversations, two different metrics were calculated. The calculation of both metrics required a series of successive steps that will be explained further.

Transcription of real-life conversations First, the conversations were consistently transcribed word for word following specified transcription rules. To analyze talk turns, transcribers had to create sequences of clean-cut speaking turns for each conversation, assigning each speaking turn to one of the two speakers. On the basis of unitizing guidelines from interaction analysis (Auld & White, 1956; Hatfield & Weider-Hatfield, 1978), we defined a statement as our unit of analysis. A unit begins when Speaker A says his/her first word, and it ends when Speaker B utters his/her first word. The speaking turns in real-life conversations are rarely as obvious as those in the *Romeo and Juliet* example provided—for example, when interlocutors talk over one another during the conversation. In such events, transcribers artificially created clean-cut sequences by following the time course of statements and transcribing them one after another. We additionally built on the guidelines for the transcription of oral language samples provided in the LIWC 2007 manual (e.g., transcription of disfluencies [hmm, uh-uh, etc.], stuttering, and transcriber's comments) (Pennebaker et al., 2007). When finalized, the transcripts comprised all statements—that is, all words and disfluencies uttered by each speaker in clean-cut talk-turns following the timely process of the conversations.

Language style analysis Second, since our analyses were based on conversations conducted in German, the German version of the LIWC dictionary was used for the language style analysis. The current version of the German LIWC dictionary contains the following function word categories: *Pronoun* [with the personal pronoun categories *I* (e.g., *I, me, mine*), *we* (e.g., *our, us, we*), *self* (e.g., *myself, us, I*), *you* (e.g., *you, thee, thine*), and *other* (e.g., *he, him, they*)], *negation*

(e.g., *no, not*), *assent* (e.g., *ok, yes*), *article* (e.g., *a, an, the*), and *preposition* (e.g., *above, at, into*).

The *pronoun* category represents the overall proportion of pronouns used in the unit of analyses and can be fragmented into the separate personal pronoun categories: *I, we, self, you,* and *other*. The individual's pronoun use is linked to different psychological aspects and changes thereof (e.g., depression or status; for an overview, see Chung & Pennebaker, 2007; Pennebaker, 2013). To the best of our knowledge, there is no work on the coordination of pronouns, especially personal pronouns, in dyadic interaction or their respective function. To consider the dictionary's hierarchical structure regarding pronouns and potential differences in pronoun use in a dyad, we calculated four mean scores: rLSM and LSM including all nine individual function word categories (*I, we, self, you, other, negate, assent, article, preposition*), and rLSM_p and LSM_p, where the individual pronoun categories *I, we, self, you,* and *other* are summarized under the *pronoun* category. Thus, the rLSM_p and LSM_p scores include the categories *pronoun, negate, assent, article,* and *preposition*.

LSM

LSM scores were calculated on the basis of the formula proposed by Ireland and Pennebaker (2010; Eq. 5). Prior to language style analyses, transcripts were separated by speaker, producing two independent documents—one containing all statements of Speaker A, the second containing all statements of Speaker B. Each document was then analyzed using the LIWC software. LIWC results for each speaker were entered into Eq. 5 for each of the function word categories, producing one score per category. On the basis of these scores, LSM and LSM_p were calculated. The LSM scores are calculated using R Studio.

rLSM

The calculation of rLSM is based on the metric introduced in this article, mathematically represented in Eq. 14. To calculate rLSM scores, first Speaker A and B are assigned on the basis of the opening statements of the conversation—the person who starts the conversation and all his/her following statements are assigned to Speaker A, whereas the second statement and all his/her following statements are assigned to Speaker B. LIWC analyses produce results for each function word category and statement of a conversation. rLSM scores are calculated according to Eq. 14, with *A* and *B* representing the speakers in the respective conversation, *C* being all available function word categories, and *S* representing all consecutive statements of each conversation. On the basis of the rLSM scores for each category, rLSM and rLSM_p are calculated. rLSM scores were calculated using R Studio.

Table 6 Results of Shapiro–Wilk tests

Category	rLSM			LSM		
	D	df	p	D	df	p
rLSM/LSM	.96	69	.032	.89	77	<.0001
rLSM_p/LSM_p	.99	75	.612	.88	77	<.0001
Pronoun	.96	77	.028	.92	77	<.0001
I	.99	77	.728	.91	77	<.0001
We	.86	71	<.0001	.88	77	<.0001
Self	.97	77	.108	.88	77	<.0001
You	.91	76	<.0001	.87	77	<.0001
Other	.88	76	<.0001	.88	77	<.0001
Negation	.98	75	.339	.82	77	<.0001
Assent	.99	77	.541	.95	77	.004
Article	.98	77	.270	.81	77	<.0001
Preposition	.99	77	.658	.93	77	<.0001

Results

Data analysis

All statistical analyses reported were performed using SPSS 24. First, we performed Shapiro–Wilk tests to test whether rLSM and LSM scores differ significantly from a normal distribution. Since results indicate a significant deviation from normality for all 12 LSM scores as well as six of the 12 rLSM scores, we used nonparametric methods to test our hypothesis. The results of Shapiro–Wilk tests can be found in Table 6.

Are rLSM values significantly lower than LSM values?

To test whether the values calculated using the rLSM metric were significantly lower than values calculated with the established LSM metric (Hypothesis 1), we performed a Wilcoxon signed-rank test for each of the 12 scores under investigation. There was a significant difference in the average scores for rLSM (Mdn = .17) and LSM (Mdn = .82), $z = -7.22$, $p < .0001$, $r = -.87$, as well as rLSM_p (Mdn = .25) and LSM_p (Mdn = .87), $z = -7.53$, $p < .0001$, $r = -.87$, with rLSM values being significantly lower than LSM values. Taking a closer look at the category level, all Wilcoxon signed-rank test for the individual language style categories revealed the same results. These results can be found in Table 6. A graphical representation of the results can be found in Fig. 1. To conclude, Hypothesis 1 was confirmed for the average rLSM and LSM scores, as well as for the individual language style categories. All effect sizes can be interpreted as large (Cohen, 1988).

Exploring the conceptual similarity of rLSM and LSM

To test for conceptual similarity between rLSM and LSM—as postulated in Hypothesis 2—12 Kendall's tau correlation coefficients were calculated. There was no relationship between the LSM and rLSM scores ($r_\tau = .09$, $p = .15$) if all personal pronoun categories were considered individually. However, if the general use of pronouns was considered, rLSM_p and LSM_P scores were significantly related ($r_\tau = .16$, $p = .023$). Similar results were found for the individual personal pronoun categories, where rLSM and LSM scores for the categories *pronoun* ($r_\tau = .03$, $p = .369$), *I* ($r_\tau = .10$, $p = .104$), *we* ($r_\tau = .04$, $p = .320$), *self* ($r_\tau = .06$, $p = .129$), and *you* ($r_\tau = -.03$, $p = .373$) were not significantly related. Only the scores for the personal pronoun category *other* ($r_\tau = .18$, $p = .009$) were significantly related. The rLSM and LSM scores for the remaining function word categories were all significantly related [*negation* ($r_\tau = .17$, $p = .017$), *assent* ($r_\tau = .10$, $p = .091$), *article* ($r_\tau = .20$, $p = .004$), and *preposition* ($r_\tau = .13$, $p = .051$)]. All effect sizes can be interpreted as small (Cohen, 1988). The results are displayed in Table 7.
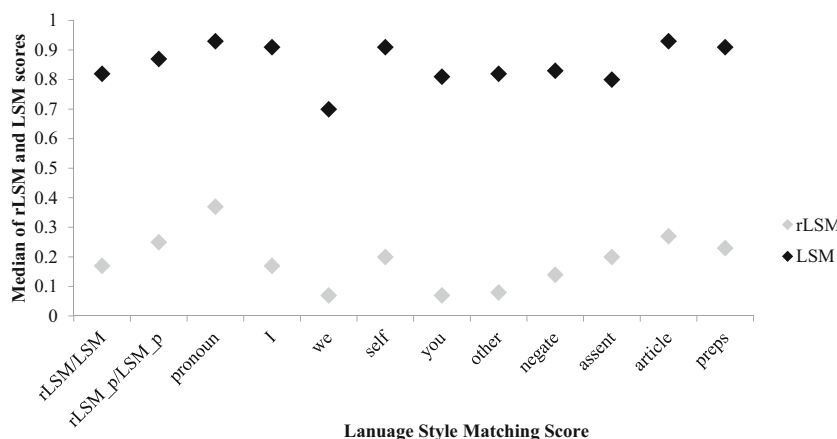
Fig. 1 Comparison of rLSM and LSM scores across all function word categories available in the German LIWC dictionary

Table 7 Results of Wilcoxon signed-rank tests and correlational analyses

Category	rLSM		LSM		n	z	r	r_τ
	M (SD)	Mdn	M (SD)	Mdn				
rLSM/LSM	.17 (.04)	.17	.81 (.07)	.82	69	− 7.22	− .87	.09
rLSM_p/LSM_p	.25 (.05)	.25	.86 (.05)	.87	75	− 7.53	− .87	.16[**]
Pronoun	.39 (.08)	.37	.92 (.06)	.93	77	− 7.62	− .87	.03
I	.17 (.08)	.17	.87 (.11)	.91	77	− 7.62	− .87	.10
We	.10 (.10)	.07	.65 (.26)	.70	71	− 7.32	− .87	.04
Self	.20 (.08)	.20	.89 (.09)	.91	77	− 7.62	− .87	.06
You	.08 (.06)	.07	.77 (.21)	.81	76	− 7.60	− .87	− .03
Other	.08 (.07)	.08	.77 (.19)	.82	76	− 7.57	− .87	.18[**]
Negation	.14 (.07)	.14	.78 (.20)	.83	75	− 7.53	− .87	.17[*]
Assent	.21 (.06)	.20	.79 (.14)	.80	77	− 7.62	− .87	.10[†]
Article	.26 (.09)	.27	.90 (.10)	.93	77	− 7.62	− .87	.20[**]
Preposition	.24 (.09)	.23	.91 (.07)	.91	77	− 7.62	− .87	.13[†]

[*] $p < .05$. [**] $p < .01$

Discussion

Researchers interested in behavior coordination have good reasons to examine LSM, because it allows them to uncover nonconscious processes of verbal behavior coordination between two conversational partners, independent of the conversational topic. Because this form of verbal behavior coordination influences various measures of relationship quality (Ireland et al., 2011; Lord et al., 2015) and can therefore be defined as an important aspect of nonconscious interpersonal behavior coordination in dyads, there is a need to standardize the conceptualization of LSM and adapt the measurement accordingly. Therefore, the primary goal of this study was to standardize the methodology in research on LSM and provide guidelines for authors planning research in this field. On the basis of the theoretical foundations of LSM, properties desirable for the analyses of LSM were derived. These properties were used to review the existing methodological approaches to LSM. This conceptual and methodological review showed that the metrics used to assess LSM so far do not fulfill the properties desirable for the adequate measurement of LSM in dynamic interaction. Hence, we developed an integrated metric that fulfills these properties and used real-life dyadic conversations to empirically test our hypothesized assumptions.

Less is more—Are rLSM scores a truer estimation of LSM than LSM scores?

The results of this study show that the most prominent LSM metric overestimates LSM scores by not considering the statement-based reciprocity between speakers in natural conversations. As a result, scores calculated with the old metric are significantly higher than scores of the same conversations assessed with the rLSM metric. These results support the conceptual framework of rLSM. With the old metric, LSM was assessed on a conversational level, that is, conversations were separated by speaker, transforming them from dynamic processes into two independent texts. By doing so, the amount of words in each of the analyzed texts increased, as compared to analyzing single statements. At the same time, the probability of finding a higher variety in function words in the separated text files increased leading to higher LSM scores. We propose that by not considering the talk-turn based similarity, the old metric rather reflects a balanced use of similar function words than a dynamic matching of language styles over time.

Even though, the magnitude of behavior coordination is difficult to compare across scholars as they all use different methodological approaches (e.g., Abney et al., 2014; Chartrand & Lakin, 2013; Lumsden et al., 2012), all scholars take the dynamic nature of coordination into account. As such, values calculated using the rLSM metric better relate to the field of research.

Since this was the first study to assess coordination using the rLSM metric, the question arises of whether rLSM is a truer estimation of coordination than LSM. To support our claim, we simulated five examples of function word distributions. All five examples comprised six statements (three talk turns) from two speakers (A and B). Because function words account for almost 60% of the words we utter (Pennebaker, 2013), we used function word scores between 0 and .6 in our data simulation, rounded to a single decimal point to facilitate understanding. The examples can be found in Table 8. Equations 5 (LSM) and 14 (rLSM) were used to calculate the rLSM and LSM scores displayed.

Table 8   Data simulation of function word use to illustrate differences in rLSM and LSM values

Statement	Function Word Score per Speaker				
	Example 1	Example 2	Example 3	Example 4	Example 5
Statement 1 – A	.60	.60	.50	.60	.60
Statement 1 – B	.60	.00	.10	.50	.10
Statement 2 – A	.60	.60	.30	.10	.30
Statement 2 – B	.60	.00	.40	.60	.30
Statement 3 – A	.60	.60	.60	.40	.10
Statement 3 – B	.60	.00	.60	.30	.60
Summary Scores					
M^{FW}_A	.60	.60	.47	.37	.33
M^{FW}_B	.60	.00	.37	.47	.33
rLSM	1.00	.00	.70	.64	.51
LSM	1.00	.00	.88	.88	1.00

A = Speaker A, B = Speaker B, M = mean, FW = function words

Examples 1 and 2 show that rLSM and LSM scores are identical as long as both speakers either use the exact same number of function words (Example 1) or one speaker does not use function words at all (Example 2). Both of these cases are very unlikely in real-life conversations. As soon as the use of function words fluctuates across different talk turns—which is most likely to happen in real-life conversations—the rLSM and LSM metrics result in very different scores.

In Examples 3 and 4, the use of function words varies randomly across talk turns. In both examples, the differences in function word use between the two speakers are identical (ΔM_{AB} = .10; Example 3: M_A = .47 and M_B = .37; Example 4: M_A = .37 and M_B = .47), resulting in identical LSM scores of LSM_{AB} = .88. The rLSM scores, on the other hand, differ from each other, with $rLSM_{AB}$ = .70 in Example 3 and $rLSM_{AB}$ = .64 in Example 4.

The most obvious example to support our claim that rLSM is a truer estimate of coordination than is LSM is Example 5: On the conversational level, both speakers use the same number of function words (M_A = .33, M_B = .33), leading to an LSM score of LSM_{AB} = 1, which represents perfect coordination. If you take a closer look at the function word scores on the talk-turn level, however, you see that the uses of function words are not identical across all subsequent statements, which is accurately captured by the much lower rLSM score of $rLSM_{AB}$ = .51.

Same same—but different?

This study further indicates conceptual similarity between rLSM and LSM if pronoun use is considered on a general level, whereas rLSM and LSM scores are not related if all personal pronoun categories are considered separately. The mostly significant, positive relationships for all function word categories-except the pronoun categories-with only low effect sizes indicate that rLSM and LSM are conceptually related but still two distinct constructs (Babcock et al., 2013), justifying the newly introduced rLSM metric. However, this study was not able to establish this conceptual similarity for all tested categories. No relationship between rLSM and LSM scores was found for the general pronoun category as well as the separate personal pronoun categories (*I*, *we*, *self*, and *you)*.

Pronouns play a special role in everyday language use. As opposed to the other function word categories, general pronoun use, as well as the use of specific personal pronouns, is related to manifold psychological variables—for example, status (Kacewicz, Pennebaker, Davis, Jeon, & Graesser, 2014), depression (Rude, Gortner, & Pennebaker, 2004), or reactions to life stressors (Pennebaker & Lay, 2002). Even though pronouns play such an important part in language use, there is no work on the coordination of (personal) pronouns in dyads and the functions thereof, to the best of our knowledge. As such, the most plausible explanation for the results is the conversational content.

The conversations analyzed in this study were aimed at initiating contact between two previously unacquainted interlocutors. Thus, the conversations were partly structured inasmuch as a basic conversational topic was provided. By following the instructions, both conversational partners had to disclose some personal information about themselves throughout the conversation. Because personal pronouns are uttered whenever a person talks about him/herself or commonalities with others (Pennebaker et al., 2015), the personal disclosure is most likely represented by the use of words that fall in the (personal) pronoun categories. In a statement-based rLSM analyses, a high rLSM score for category *I*, for example, could result in Speaker A speaking about his/her family in detail, and Speaker B directly responding with a report of his/

her own family in the consecutive statement. In a different type of conversation—like the conversations analyzed in this study—Speaker A could talk about his/her family, and Speaker B could, in turn, ask questions about it or comment on Speaker A's statements, resulting in low rLSM values for this pair of statements. As such, high LSM values indicate a high similarity in pronoun use across the conversation, representing similar levels of self-disclosure, whereas the non-significant relationships between both scores and the lower rLSM values indicate different patterns of pronoun use on the talk-turn level. Instead, interview-like conversations, with Speaker A talking about his/her personal details and Speaker B asking questions about these details, could occur. Then, after some time over which the most important aspects are covered, the speakers switch roles, with Speaker B talking about his/her personal details and Speaker A asking questions, resulting in rather high LSM values in these categories across the whole conversation, and rather low rLSM values at the same time.

Previous work on role assignment supports our explanation by showing that role assignment sensitive to the contextual demands of a task may be linked to better performance in the task (Abney et al., 2014). Hence, a low correlation between the LSM and rLSM values in these categories does not necessarily speak against conceptual similarity, but could rather indicate different patterns of pronoun use at the talk-turn and conversational levels.

Where do we go from here?

On the basis of the results at hand, new directions for future research on LSM emerge. We were able to reconceptualize LSM and to derive an appropriate metric based on existing research. Applied to real-life conversations, our theoretical and methodological assumptions were mostly confirmed. With rLSM and LSM metrics being related but not identical, future research will need to further investigate the relationship between the new rLSM metric and well-established outcomes of social interaction—for example, relationship initiation and stability (Ireland et al., 2011), empathy (Lord et al., 2015), or perceived social support (Rains, 2015). Additionally, benchmarks for high and low LSM need to be redefined. The current benchmarks for low (LSM = .60) and high (LSM = .85) LSM (Cannava & Bodie, 2016) are not adequate when using the rLSM metric.

Furthermore, we did not explore the whole potential the new rLSM metric offers: It is not only able to measure LSM as a dyadic, multi-dimensional, dynamic construct, but further captures the directional adaption to the respective other. By analyzing individual and talk-turn based LSM, leader-follower interactions in dyads can be uncovered and applied to basic social interactions in experimental contexts as well as applied dyadic settings in future studies. Moreover, the rLSM

metric offers the potential to analyze the whole process of adaption by looking at the development over time. One possible application is the examination of theoretically based phases of a conversation (Meinecke, Klonek, & Kauffeld, 2016), or the development of LSM over time.

Because the rLSM metric itself is atheoretical, in that it captures coordination without presupposing the effects of coordination, applying rLSM to various dyadic contexts could also support disentangling the discordant theoretical views on the automatic and context-sensitive nature of linguistic behavior coordination (e.g., Fusaroli et al., 2014; Riley et al., 2011). The phenomenon of LSM has not been studied in dyadic research exclusively, but has also gained attention in group and team settings (e.g., Gonzales, Hancock, & Pennebaker, 2010; Yilmaz, 2016). Therefore, the rLSM metric needs to be adapted to the team context. If the metric is applied to the team context, only the number of interlocutors and the potential time lag of adaption will need to be altered—all other conceptual and methodological underpinnings will remain identical to the description in this study.

Guidelines for authors

One of the main aims of this study was to provide guidelines for future authors on the use of the different LSM metrics based on theoretical and methodological underpinnings and from the empirical results of this study. An overview of the guidelines can be found in Table 4. We propose that the metric most commonly used for the assessment of LSM, introduced by Ireland and Pennebaker (2010), captures the similarity of language styles in at least two static texts. Under static texts, we summarize all forms of written texts that are self-contained and function without the direct reaction of an interactional partner—for example, letters, prose, or transcripts of speeches. The study that introduced the metric also offers excellent examples of its application (Ireland & Pennebaker, 2010).

Furthermore, we reason that, when applied to dynamic texts, the Ireland and Pennebaker (2010) metric captures a balanced use of function words between two conversational partners rather than dynamic coordination. Hence, we propose that as soon as any kind of conversational dependency is assumed, the new rLSM metric will need to be applied, since its application in a dyadic context was empirically confirmed in this article.

The metric introduced by Danescu-Niculescu-Mizil et al. (2011) for the analysis of language style accommodation in written online conversations is most properly used in the context it was developed for—that is, analyses of big data on social media platforms such as Twitter. Even though conceptual overlaps between LSM and the metric by Danescu-Niculescu-Mizil et al. have been established, future studies should examine the statistical similarity of these metrics

before applying the latter to the context of natural conversations. For now, we can only recommend its application to the designated context if the context is labeled appropriately.

Conclusion

Even though the definition of LSM varies, analyses of the phenomenon and its influence on social interactions form a relevant field of research that has gained more and more attention in the last several years. This study provides common ground for future research by reviewing the methodological approaches to this field of research and introducing an integrated metric covering the gap between the conceptualization and methodology of this phenomenon. Our new metric needs to be applied to different contexts to further establish its validity, but it offers great potential to standardize the research on LSM in the future.

References

Abney, D. H., Paxton, A., Dale, R., & Kello, C. T., (2014). General complexity matching in dyadic conversation. *Journal of Experimental Social Psychology*, *143*, 2304–2315. https://doi.org/10.1037/xge0000021

Auld, F., & White, A. M. (1956). Rules for dividing interviews into sentences. *Journal of Psychology*, *42*, 273–281.

Babcock, M. J., Ta, V. P., & Ickes, W. (2013). Latent semantic similarity and language style matching in initial dyadic interactions. *Journal of Language and Social Psychology*, *33*, 78–88. https://doi.org/10.1177/0261927X13499331

Bernieri, F. J., & Rosenthal, R. (1991). Interpersonal coordination: Behavior matching and interactional synchrony. In R. S. Feldman & B. Rimé (Eds.), Fundamentals of nonverbal behavior: Studies in emotion & social interaction (pp. 401–432). New York: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.004

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Bowen, J. D., Winczewski, L. A., & Collins, N. L. (2017). Language style matching in romantic partners: Conflict and support interactions. *Journal of Language and Social Psychology*, *36*, 263–286. https://doi.org/10.1177/0261927X16666308

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, *1*, 274–291. https://doi.org/10.1111/j.1756-8765.2009.01019.x

Cannava, K., & Bodie, G. D. (2016). Language use and style matching in supportive conversations between strangers and friends. *Journal of Personal and Social Relationships*, *34*, 1–19. https://doi.org/10.1177/0265407516641222

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*, 893–910. https://doi.org/10.1037/0022-3514.76.6.893

Chartrand, T. L., & Lakin, J. L. (2013). The antecedents and consequences of human behavioral mimicry. *Annual Review of Psychology*, *64*, 285–308. https://doi.org/10.1146/annurev-psych-113011-143754

Chung, C., & Pennebaker, J. (2007). The psychological functions of function words. In K. Fiedler (Ed.), Social communication (pp. 343–359). Hove: Psychology Press. https://doi.org/10.4324/9780203837702

Clark, H. H. (1996). Using language. New York: Cambridge University Press.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale: Erlbaum.

Danescu-Niculescu-Mizil, C., Gamon, M., & Dumais, S. (2011). Mark my words! Linguistic style accommodation in social media. In *Proceedings of the 20th International Conference on the World Wide Web* (pp. 745–754). https://doi.org/10.1145/1963405.1963509

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological Science*, *23*, 931–939. https://doi.org/10.1177/0956797612436816

Fusaroli, R., Raczaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, *32*, 147–157. https://doi.org/10.1016/j.newideapsych.2013.03.005

Giles, H., & Coupland, N. (1991). Language contexts and consequences. Milton Keynes: Open University Press.

Gonzales, A. L., Hancock, J. T., & Pennebaker, J. W. (2010). Language style matching as a predictor of social dynamics in small groups. *Communication Research*, *37*, 3–19. https://doi.org/10.1177/0093650209351468

Guastello, S. J., & Peressini, A. F. (2017). Development of a synchronization coefficient for biosocial interactions in groups and teams. *Small Group Research*, *48*, 3–33. https://doi.org/10.1177/1046496416675225

Hatfield, J. D., & Weider-Hatfield, D. (1978). The comparative utility of three types of behavioral units for interaction analysis. *Communication Monographs*, *45*, 44–50.

Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation, *Social Cognition*, *27*, 949–961. https://doi.org/10.1521/soco.2009.27.6.949

Ireland, M. E., & Henderson, M. D. (2014). Language style matching, engagement, and impasse in negotiations. *Negotiation and Conflict Management Research*, *7*, 1–16. https://doi.org/10.1111/ncmr.12025

Ireland, M. E., & Pennebaker, J. W. (2010). Language style matching in writing: Synchrony in essays, correspondence, and poetry. *Journal of Personality and Social Psychology*, *99*, 549–571. https://doi.org/10.1037/a0020386

Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychological Science*, *22*, 39–44. https://doi.org/10.1177/0956797610392928

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, *33*, 125–143. https://doi.org/10.1177/0261927X13502654

Klonek, F. E., Quera, V., & Kauffeld, S. (2015). Coding interactions in Motivational Interviewing with computer-software: What are the advantages for process researchers? *Computers in Human Behavior*, *44*, 284–292. https://doi.org/10.1016/j.chb.2014.10.034

Lakin, J. L., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science*, *14*, 334–339. https://doi.org/10.1111/1467-9280.14481

Lord, S. P., Sheng, E., Imel, Z. E., Baer, J., & Atkins, D. C. (2015). More than reflections: Empathy in motivational interviewing includes language style synchrony between therapist and client. *Behavior Therapy*, *46*, 296–303. https://doi.org/10.1016/j.beth.2014.11.002

Lumsden, J., Miles, L. K., Richardson, M. J., Smith, C. A., & Macrae, C. N. (2012). Who syncs? Social motives and interpersonal coordination. *Journal of Experimental Social Psychology*, *48*, 746–751. https://doi.org/10.1016/j.jesp.2011.12.007

Main, A., Paxton, A., & Dale, R. (2016). An exploratory analysis of affective dynamics between mothers and adolescents during conflict discussions. *Emotion*, *16*, 913–928. https://doi.org/10.1037/emo0000180

Meinecke, A. L., Klonek, F. E., & Kauffeld, S. (2016). Appraisal participation and perceived voice in annual appraisal interviews: Uncovering contextual factors. *Journal of Leadership and Organizational Studies*, *24*, 230–245. https://doi.org/10.1177/1548051816655990

Meyer, A. S., & Bock, K. (1999). Representations and processes in the production of pronouns: Some perspectives from Dutch. *Journal of Memory and Language*, *41*, 281–301. https://doi.org/10.1006/jmla.1999.2649

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). *Distributed representations of words and phrases and their compositionality.* arXiv:1310.4546v1

Muir, K., Joinson, A., Cotterill, R., & Dewdney, N. (2016). Characterizing the linguistic chameleon: Personal and social correlates of linguistic style accommodation. *Human Communication Research*, *24*, 462–484. https://doi.org/10.1111/hcre.12083

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, *21*, 337–360. https://doi.org/10.1177/026192702237953

Paxton, A., & Dale, R. (2013). Frame-differencing methods for measuring bodily synchrony in conversation. *Behavior Research Methods*, *45*, 329–343. https://doi.org/10.3758/s13428-012-0249-2

Paxton, A., Dale, R., & Richardson, D. C. (2016). Social coordination of verbal and nonverbal behaviours. In P. Passos, K. Davis, & J. Y. Chow (Eds.), Interpersonal coordination and performance in social systems (pp. 259–274). Abingdon: Routledge.

Pennebaker, J. W. (2013). The secret life of pronouns—What our words say about us. London: Bloomsbury.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). Operator's manual: Linguistic inquiry and word count—LIWC2007. Austin: University of Texas, Department of Psychology. Retrieved from liwc.net

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. Austin: University of Texas, Department of Psychology. https://doi.org/10.1068/d010163

Pennebaker, J. W., & Lay, T. C. (2002). Language use and personality during crises: Analyses of mayor Rudolph Giuliani's press conferences. *Journal of Research in Personality*, *36*, 271–282. https://doi.org/10.1006/jrpe.2002.2349

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, *54*, 547–577. https://doi.org/10.1146/annurev.psych.54.101601.145041

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169–190, disc. 190–226. https://doi.org/10.1017/S0140525X04000056

Rains, S. A. (2015). Language style matching as a predictor of perceived social support in computer-mediated interaction among individuals coping with illness. *Communication Research*, *43*, 694–712. https://doi.org/10.1177/0093650214565920

Riley, M. A., Richardson, M. J., Shockley, K., & Ramenzoni, V. C. (2011). Interpersonal synergies. *Frontiers in Psychology*, *2*, 38:1–7. https://doi.org/10.3389/fpsyg.2011.00038

Rude, S. S., Gortner, E. M., & Pennebaker, J. W. (2004). Language use of depressed and depression-vulnerable college students. *Cognition and Emotion*, *18*, 1121–1133.

Segalowitz, S. J., & Lane, K. (2004). Perceptual fluency and lexical access for function versus content words. *Behavioral and Brain Sciences*, *27*, 307–308. https://doi.org/10.1017/S0140525X04310071

Shakespeare, W. (1597). Romeo and Juliet. Derived from https://www.opensourceshakespeare.org/views/plays/play_view.php?WorkID=romeojuliet&Act=1&Scene=1&Scope=scene

Shephard, C., Giles, H., & LePoire, B. (2001). Communication accommodation theory. In W. P. Robinson & H. Giles (Eds.), The new handbook of language and social psychology (pp. 33–56). Chichester: Wiley. https://doi.org/10.4135/9781483329529.n12

Taylor, P., & Thomas, S. (2008). Linguistic style matching and negotiation outcome. *Negotiation and Conflict Management*, *1*, 1–34. doi: https://doi.org/10.1111/j.1750-4716.2008.00016.x

Tylén, K., Fusaroli, R., Bundgaard, P. F., & Østergaard, S. (2013). Making sense together: A dynamical account of linguistic meaning-making. *Semiotica*, *194*, 39–62. https://doi.org/10.1515/sem-2013-0021

Yilmaz, G. (2016). What you do and how you speak matter: behavioral and linguistic determinants of performance in virtual teams. *Journal of Language and Social Psychology*, *35*, 76–97. https://doi.org/10.1177/0261927X15575772