Bilingualism: Language and Cognition

cambridge.org/bil

Complexity matching and lexical matching in monolingual and bilingual conversations

Sara Schneider, Adolfo G. Ramirez-Aristizabal, Carol Gavilan and Christopher T. Kello

Cognitive and Information Sciences, University of California, Merced

Research Article

Cite this article: Schneider S, Ramirez-Aristizabal AG, Gavilan C, Kello CT (2020). Complexity matching and lexical matching in monolingual and bilingual conversations. *Bilingualism: Language and Cognition* **23**, 845–857. https://doi.org/10.1017/S1366728919000774

Received: 18 September 2018
Revised: 10 October 2019
Accepted: 31 October 2019
First published online: 9 December 2019

Keywords:
bilingual conversation; complexity matching; lexical matching; convergence; alignment; hierarchical temporal structure

Address for correspondence: Sara Schneider, E-mail: sschneider2@ucmerced.edu.

Abstract

When people interact, aspects of their speech and language patterns often converge in interactions involving one or more languages. Most studies of speech convergence in conversations have examined monolingual interactions, whereas most studies of bilingual speech convergence have examined spoken responses to prompts. However, it is not uncommon in multilingual communities to converse in two languages, where each speaker primarily produces only one of the two languages. The present study examined complexity matching and lexical matching as two measures of speech convergence in conversations spoken in English, Spanish, or both languages. Complexity matching measured convergence in the hierarchical timing of speech, and lexical matching measured convergence in the frequency distributions of lemmas produced. Both types of matching were found equally in all three language conditions. Taken together, the results indicate that convergence is robust to monolingual and bilingual interactions because it stems from basic mechanisms of coordination and communication.

Introduction

Researchers estimate that more than half of the global population speaks more than one language (i.e., Bialystok, Craik & Luk, 2012; Grosjean, 2010; Romaine, 2012), and when learning starts at an early age, native or near-native proficiency in comprehension and production in both first and second languages is commonly achieved (Perani, Abutalebi, Paulesu, Brambati, Scifo, Cappa & Fazio, 2003). Proficient bilingual and multilingual speakers switch between languages in certain conversational contexts, with little apparent difficulty (Fricke & Kootstra, 2016; Toribio, 2004). Sometimes the use of each language is asymmetric between speakers. For instance, in conversations between older and younger members of immigrant families, the elders may prefer to speak their heritage language whereas the younger generation may speak the language of their new community (e.g., Park & Sarkar, 2007). Communicating in two different languages concurrently is a relatively normal mode of bilingual communication known as LINGUA RECEPTIVA (Bahtina-Jantsikene & Backus, 2016; Bahtina, ten Thije & Wijnen, 2013; ten Thije, Gooskens, Daems, Cornips & Smits, 2017; ten Thije, 2013).

The prevalence of lingua receptiva raises the question of whether principles of language interaction in monolingual speech may apply to bilingual interactions as well. One of the most well-established principles of monolingual interaction is SPEECH ALIGNMENT (Pickering & Garrod, 2004), which generally refers to the matching of speech produced with speech perceived across multiple levels of representation. For instance, in a conversation between two people, one speaker may use the word "loafer" to refer to a shoe, and the other person may then use the same word "loafer" instead of, say, a "dress shoe" (Brennan & Clark, 1996). Convergence can also occur in the phonetic features of speech (Kim, Horton & Bradlow, 2011; Pardo, 2006), syntactic structures (Bock, 1986), and prosody (Abney, Paxton, Dale & Kello, 2014; De Looze, Scherer, Vaughan & Campbell, 2014; Xia, Levitan & Hirschberg, 2014).

In the present study, we investigate speech convergence in open-ended conversations using two measures of matching that can be applied within one language, as well as across two different languages. Conversations were spoken in English, Spanish, or in a "Mixed" condition where one person spoke English and the other Spanish. The study of *inter-language* speech convergence in conversations is challenging because there may be no direct correspondences between the surface forms of linguistic units or features produced in two different languages, like English and Spanish. We use COMPLEXITY MATCHING (Abney, Kello & Warlaumont, 2015; Abney et al., 2014) as a recent measure of speech convergence that can be directly applied to acoustic speech signals without linguistic coding. It is therefore equally applicable to measuring convergence of speech signals in the same or different languages. We also use a measure of LEXICAL MATCHING (Brennan & Clark, 1996; Brennan, Kuhlen & Charoy, 2018; Garrod & Anderson, 1987; Niederhoffer & Pennebaker, 2002) that can be applied to open-ended conversations without one-to-one alignment of words, although it does require some degree of translation for estimating semantic correspondences of lemmas used across languages.

CAMBRIDGE UNIVERSITY PRESS

Our goal was to measure matching within one language and compare it with matching across two languages, using an acoustic measure of the speech signals produced, compared with a linguistic measure of the underlying language interaction. Prior studies lead us to predict that both kinds of matching should occur in purely Spanish conversations and purely English conversations, provided that speakers are proficient enough with the language. It is an open question whether conversations across languages might exhibit weaker signs of matching due to translation, or whether matching is more basic to spoken interactions as a form of convergence during communicative coordination.

In the next section, we review prior studies of inter- and intra-language convergence, pointing out the need for measures spanning different levels of analysis that may be applied more generally to open-ended conversations. We then introduce and explain complexity matching and our measure of lexical matching, where the latter is based on overlap in the frequency distributions of lemmas underlying word tokens. Our experiment follows, and we end with a discussion of the implications of our results for theories of bilingualism, language interaction, and convergence.

Monolingual and bilingual speech convergence

Whether speaking in either inter- or intra-language settings, conversations are coordinated interactions where speakers variably lead, follow, and echo each other as they exchange ideas. Convergent coordination has been observed in multiple aspects of speech and language. In terms of sound, phonetic features, such as vowel quality and voice onset time, become more similar between conversational partners. Phonetic convergence has been observed in monolingual conversations (Pardo, 2013; Pardo, 2006; Pardo, Gibbons, Suppes & Krauss, 2012) and has been indirectly explored in at least one bilingual study as well (Sancier & Fowler, 1997). Phonetic convergence in monolingual conversations was measured by Pardo (2006) and Pardo et al. (2012) by having listeners judge the similarity of speakers' phoneme production. Using a more indirect method, Nielsen (2011) demonstrated that phonetic convergence depends on factors such as word frequency and voice onset time (VOT).

Building on Nielsen (2011), convergence in VOT has also been used to measure bilingual phonetic imitation (Balukas & Koops, 2015; Tobin, Nam & Fowler, 2017). For example, Balukas and Koops (2015) analyzed words spoken in conversational interviews by New Mexico Spanish–English bilinguals. The words analyzed contained an initial /p/, /t/, or /k/ sound in both Spanish and English. Participants speaking Spanish (the first language for most participants) were found to have VOT values within the normal range for monolingual Spanish. However, when speaking English, VOTs fell within the low range of monolingual English. Therefore, participants appeared to "bend" phonemes of their non-dominant language towards the dominant language near code-switching events.

At more grammatical levels of language processing, syntactic priming is a form of conversational coordination where speakers tend to produce (Bock, 1986; Healey, Purver & Howes, 2014) or comprehend (Branigan, Pickering, Liversedge, Stewart & Urbach, 1995) new sentences using syntactic structures recently produced or heard. Syntactic priming is well-established in monolingual conversations (Hardy, Messenger & Maylor, 2017), and bilingual conversations as well because languages often share common syntactic constructions (Hartsuiker, Pickering & Veltkamp, 2004), making it relatively easy to apply measures of syntactic priming in both monolingual and bilingual studies.

While researchers have found evidence for syntactic priming across languages, the language tasks used are often contrived and sometimes require confederates (Fleischer, Pickering & McLean, 2012; Hartsuiker et al., 2004). For instance, Hartsuiker et al. (2004) had Spanish–English bilingual participants talk about cards in English with a bilingual confederate who spoke in Spanish. Participants who heard a passive sentence in Spanish were more likely to respond using a passive sentence in English. The authors concluded that these findings provide support for the integration of syntactic representations between the two languages. Kantola and van Gompel (2011) also found syntactic priming in Swedish–English bilinguals, and recent corpus-based studies on bilingual syntactic priming have provided a more naturalistic source of evidence (Gries & Koostra, 2017).

Speech convergence has been theorized to be beneficial to language interactions by helping to establish common ground (Brennan & Clark, 1996), affiliation (Manson, Bryant, Gervais & Kline, 2013; Pardo et al., 2012), or comprehension (Branigan et al., 1995; Schober & Clark, 1989). Convergence may stem from domain-general processes of imitation (De Looze, Oertel, Rauzy & Campbell, 2011; van Baaren, Holland, Steenaert & van Knippenberg, 2003), possibly implemented through links between speech and language perception and production (Buchsbaum, Gregory & Colin, 2001; Tian & Poeppel, 2012), or on-the-fly processes and representations that arise to support coordination and understanding (Brennan & Hanna, 2009).

Findings of convergence in bilingual speakers suggest that similar mechanisms of matching are at play in both bilingual and monolingual conversations (Fricke & Kootstra, 2016; Kroll, Dussias, Bogulski & Kroff, 2012), which may or may not be symmetric between two given languages spoken (Blumenfeld & Marian, 2007). The rationale for similar mechanisms can be explained in terms of the theory of interactive alignment (Garrod & Pickering, 2004; Pickering & Garrod, 2004) in which convergence stems from the alignment of language representations across multiple levels of processing that interact with each other. To the extent that convergence is found across two languages, interactive alignment entails that representations from different languages must be aligned. Representational structures may have direct correspondences across languages in some cases, such as shared discourse processes stemming from one culture with multiple languages. In other cases, alignment may require some degree of translation, as between Spanish and English lexicons. Costa, Pickering, and Sorace (2008) note that convergence may be weaker to the extent that speakers are less proficient in the language spoken, and one might imagine a similar weakening if speakers need to maintain activation of two languages simultaneously during a conversation.

The present study further investigates inter-language speech convergence by adopting and applying COMPLEXITY MATCHING and LEXICAL MATCHING to both intra- and inter-language conversations. In the next two sections, we explain these two measures and how they can be applied within and across languages.

Complexity matching in the nested clustering of speech events

COMPLEXITY MATCHING is a recently hypothesized expression of speech convergence and interpersonal coordination inspired by theoretical work on interactions between complex networks (Abney et al., 2014; Marmelat & Delignières, 2012; West,

Geneston & Grigolini, 2008). West et al. (2008) theorized that when complex networks interact, the exchange of information will be maximized when the distribution of events between networks share a common dynamic. Exchange of information was defined in terms of mutual impact on each other's event dynamics, and dynamics were defined in terms of the timing of network events.

Inspired by West and colleagues, Abney et al. (2014) proposed complexity matching as a measure of coordination dynamics in speech. They converted speech waveforms into time series of speech events that capture hierarchical temporal structure (HTS) in the nested clustering of acoustic speech energy. The timescale of small clusters roughly corresponds with phonemes, larger timescales with syllables and words, still larger phrases and sentences, up to the largest timescales that cover conversational turns and other discourse-level dynamics. Clustering across timescales reflects the hierarchical nesting of linguistic units such as phonemes within syllables, syllables within words, words within phrases, and so on (for direct evidence of this claim, see Falk & Kello, 2017).

Abney et al. (2014) analyzed matching in the nested relations among temporal clusters of acoustic speech events in two different types of conversations. Each pair of partners was cued to have one affiliative conversation about a topic of shared interest, and one argumentative conversation about a topic on which partners took opposing sides. Complexity matching was found in the affiliative conversation but not the argumentative conversation, indicating that the measure of nested clustering reflects more than just turn taking or other general effects of speaking together. Abney, Warlaumont, Oller, Wallot, and Kello (2017) followed up with another effect on complexity matching, this time in the vocal interactions between infants and adult caregivers. More matching was found when infants produced speech-related sounds as opposed to non-speech sounds, and adults adjusted their HTS to match that of infants, rather than vice versa.

These two prior studies provide evidence that complexity matching is not just a basic consequence of turn-taking or other general factors of speaking together – complexity matching is more prevalent during positive, communicative interactions, which is evidence that complexity matching reflects a functional convergence in speech interactions. There is also evidence that the nested clustering of events, i.e., the statistic that converges in complexity matching also reflects functional aspects of speech. Kello, Dalla Bella, Médé, and Balasubramaniam (2017) measured HTS in conversations as well as TED talk monologues spoken in six different languages, including English and Spanish. Language had no effect on HTS, but conversations yielded greater HTS than monologues, and original TED talks yielded greater HTS than speech synthesized versions of TED talk transcriptions. These effects indicate that HTS varies with discourse processes (dialog versus monologue) and prosody (natural versus synthetic), suggesting that complexity matching may vary depending on the timescales analyzed.

Lexical matching in word frequency distributions

Whereas complexity matching is a measure of convergence based on a physical measure of the speech signal (the amplitude envelope), lexical entrainment is a measure of convergence based on non-physical representations of words or lemmas. Specifically, lexical entrainment is the tendency for speakers to choose and repeat certain words as referents during conversations (Anderson, Garrod & Sanford, 1983; Brennan & Clark, 1996; Clark & Brennan, 1991). For instance, in the study by Brennan and Clark (1996) mentioned earlier, speakers were observed to form "conceptual pacts" by converging on certain words to signal certain referents. To measure lexical entrainment, the authors computed the probability of target words being produced when cued from trial to trial.

Lexical entrainment has also been found in more open-ended speech exchanges. For instance, Nenkova, Gravano, and Hirschberg (2008) measured lexical entrainment in conversations by quantifying similarities in the proportions of times conversational partners used certain words. Likewise, Levitan, Benus, Gravano, and Hirschberg (2015) measured the entrainment of turn-taking behaviors between speakers using *Kullback-Leibler Divergence* (KLD), which measures the degree to which one probability distribution is contained within another. Levitan et al. (2015) compared KLD between conversational partners and surrogate pairs. As expected, KLD was less for partners compared with surrogates, which means that the distributions of word usage for two given partners had more overlap compared with surrogates.

It is not straightforward to apply the above methods to measure lexical entrainment across languages because there may not be a clear mapping between their respective lexicons. In one study, Ni Eochaidh (2010) found lexical entrainment across languages by using a highly constrained bilingual speech task that allowed for unambiguous mappings between lexical items across English and Irish. Otherwise there is a dearth of empirical studies testing lexical entrainment across languages, although a study by Bortfeld and Brennan (1997) suggests that less facility with a second language may not interfere with the effect – they found lexical entrainment to occur equally for language interactions in which speakers were more or less proficient at the language spoken.

We use the term "lexical matching" to refer to our measure of convergence in the frequencies of lemma usage, where lemmas abstract over the surface forms of words, and provide a more consistent basis for translation across languages. We use a variant of KLD known as *Jensen-Shannon Divergence* (JSD), which is simply a symmetric version of KLD. JSD is normalized so that zero means identical probability distributions, and one means completely non-overlapping distributions, i.e., no lemmas shared by conversational partners. JSD requires a correspondence between words spoken, which is mostly a given when conversational partners are speaking the same language. When two different languages are spoken, as in lingua receptiva, we can measure overlap by translating the lemmas of one language into those of the other. Translations can vary depending on the translator, but this issue is addressed by using surrogates as baselines from the same translator. Lexical matching was measured as significant differences from a baseline for both intra- and inter-language conversations. Lexical matching is based on direct correspondences in the intra-language conditions, and translations based on corresponding semantics in the inter-language conditions.

Current experiment

In the present study, we investigated speech convergence within and between English and Spanish in naturalistic conversations using complexity matching and lexical matching as two different measures of convergence. The monolingual English condition served as a baseline and test for replicating prior findings of complexity matching and lexical matching, whereas the monolingual Spanish and bilingual Mixed conditions extended previous

investigations of complexity matching and lexical matching to a new language and to inter-language bilingual conversations. We compared the degrees of matching across language conditions, and tested whether inter-language convergence occurs in acoustic and linguistic measures of matching. The range of language backgrounds was relatively narrow in our participant pool, but we also examined whether Spanish and English, as a dominant or non-dominant language, had an effect on convergence when Spanish was spoken. Finally, we tested whether degrees of complexity and lexical matching were positively correlated with each other on the hypothesis that they share a common basis, e.g., via interactive alignment across levels of processing.

Methods

Participants

Sixty participants (mean age = 19.45; males = 8, females = 52) were recruited from the University of California, Merced, through the SONA participant pool for course credit. Participants were recruited in pairs, and two pairs were omitted from all analyses due to technical difficulties with the audio recordings (remaining subjects: mean age = 19.35; males = 5, females = 51). Our sample size is comparable to that of prior studies involving convergence in dyadic speech (Abney et al., 2014; Falk & Kello, 2017; Marmelat & Delignières, 2012; Pardo et al., 2012; Pardo, Urmanache, Gash, Wiener, Mason, Wilman, Francis & Decker, 2018). All but three dyads reported not knowing one another prior to the experiment, and the remaining three dyads were acquaintances. Participants reported their native language as either Spanish ($n = 24$), English ($n = 5$), or both Spanish and English ($n = 17$). One participant also reported Punjabi as their second language. Participants rated which language they used most comfortably on a daily basis: 7 reported Spanish, 30 English, and 14 both Spanish and English equally (5 non-responses). Finally, the native countries of origin included the United States ($n = 39$), Mexico ($n = 13$), El Salvador ($n = 2$), and both Mexico and the United States ($n = 2$). One of the native Spanish speaking experimenters listened to the conversations and confirmed that all participants were able to hold a conversation in Spanish. Additional information about the participant's average proficiency may be found in Table 1.

Table 1. Mean proficiency ratings (with standard deviations in parentheses) for English and Spanish. Percent frequency of use refers to how often each language is used per week, such that both languages could be used every day (100%) in a given week. Non-dominant language corresponded to the "second" language as asked on the survey (one participant whose second language was Punjabi was omitted from this table). The self-reported reading, writing, and speaking proficiency scores were rated out of 10, where 10 indicated total fluency.

	English	Spanish
Age of acquisition	4.5 (3.2)	1.0 (1.6)
Frequency of use for dominant language	88.2% (20)	77% (24.2)
Frequency of use for non-dominant language	81.8% (25.6)	61.8% (23.6)
Reading proficiency	9.0 (1.2)	8.2 (1.5)
Writing proficiency	8.6 (1.4)	7.1 (1.9)
Speaking proficiency	9.3 (1.1)	8.6 (1.4)

Procedure

Dyads were seated at a table in a small room (8.5' by 7') facing one another, approximately 2.5 feet apart, where they engaged in three conversations. Each conversation was recorded using an M-Audio Mobile Pre-amp, two Shure SM10A headset microphones, and the Audacity 2.0.2 audio software.

Upon arrival, participants were instructed to read and sign a consent form, which explained that they could end the experiment at any time without penalty. Participants were then instructed to turn their cell phones off or onto silent mode to avoid interruptions. Participants filled out language background and proficiency information before the study began (see Appendix A). Each pair of participants was informed that they would be engaging in three five-minute conversations with each other. The prompted topic of one conversation was movies, another music, and the third television. Finally, one of the conversations was spoken in English, one in Spanish, and the other in a Mixed condition using both languages, with language in the latter randomly assigned to speakers. Conversation topic, language condition, and order were all randomized such that each possible combination occurred with equal probability (conditions were not counterbalanced).

Dyads were prompted to introduce themselves and chat casually for a few minutes before starting the experiment, while the experimenter tested the audio. This initial conversation was in English, during which the experimenter adjusted the input gain on each microphone relative to each person's speaking voice. During each recorded conversation, pieces of paper were taped on the wall to display the current language(s) and topic of conversation, as a reminder to participants. Participants began each conversation after the researcher said 'begin' or 'start'. The researcher sat in the same room facing away from the participants to monitor the recordings and did not engage with the participants during each five-minute conversation, or trial. Upon completion of the three conversations, participants rated each conversation in terms of ease of communication and comfortability on a scale of one to five, with five being easiest and most comfortable (see Appendix B).

Each conversation was recorded to an uncompressed stereo WAV file, with the output of one microphone sent to the left channel, and other to the right channel. The input gain level for each participant was adjusted to ensure adequate recording levels while minimizing cross-talk between the microphones. Despite best efforts, a small amount of cross-talk occurred at some points during some of the recordings. Audacity was used to remove cross-talk from the recordings before analysis. A noise profile was selected based on cross-talk examples chosen manually from a visual display of the speech waveform (the experimenter listened to confirm cross-talk), and the selected profile was applied to the whole recording to filter out acoustic energy that resembled cross-talk. For this filtering function, the sensitivity parameter was always set to 25 and the frequency smoothing parameter to 3. In addition to acoustic analyses, each recording was transcribed using TranscribeMe (http://transcribeme.com) for both Spanish and English. Two researchers reviewed all the transcriptions for quality control. Researchers confirmed that as instructed, on average, dyads in the Mixed condition only accidentally switched languages an average of about two times, and they corrected themselves shortly thereafter. In the Spanish

condition, five dyads made one mistake each, and no participants made any mistakes in the English condition. No participants switched languages accidentally in the pure English or Spanish conditions.

Complexity matching and Allan Factor analysis

HTS was measured for each speaker in terms of the amount of nested clustering in peak amplitude speech events across timescales. Specifically, the following operations were performed on the waveform for each speaker (for more detail, see Kello et al., 2017): 1) peaks of the amplitude envelope were selected using a variable amplitude set to hold the number of peaks constant across speakers; 2) a log-log Allan Factor function was computed for each event series, which quantifies the change in clustering across timescales; and 3) the rate of increase in clustering was quantified in the shorter and longer timescales by fitting a regression line to each half of the AF function. Correlations in partner slopes for the longer timescales served as our main measure of complexity matching, the timescales where variations in prosody and discourse are expressed. Using correlations to measure complexity matching obviated the need for surrogate analyses because correlations have an inherent baseline of 0 for no linear relationship.

Acoustic event analysis was based on prior studies of speech event series (Abney et al., 2014, 2017; Falk & Kello, 2017; Luque, Luque & Lacasa, 2015). Events were peaks in the amplitude envelope of the speech waveform downsampled to 11 KHz, and clusters of peaks roughly correspond to units of speech like syllables, words, and phrases (Falk & Kello, 2017). As in Kello et al. (2017), peak events were identified in the amplitude envelope using two thresholds to ensure peaks were locally maximal and relatively large in amplitude. These thresholds helped filter out noise, normalize recording levels, and produce a sparse event series with the possibility of a large dynamic range in cluster sizes. Kello et al. (2017) showed that AF analyses are not substantially affected by moderate changes to these threshold settings.

Upon completion of the acoustic event analysis, nested clustering of peak amplitude events was quantified using AF analysis, which has been used previously for measuring clustering in neuronal spike events (Lowen & Teich, 1996). The event series are windowed at multiple timescales, and events are counted within each window. The difference in event counts between adjacent windows is used in a statistical measure of temporally local variance in event counts, where greater variance corresponds with greater clustering. The average degree of clustering is quantified as AF variance $A(T)$ for a given window size T (see Figure 1). $A(T)$ was computed for 11 timescales ranging from approximately 30 ms to 30 sec.

If events are clustered across timescales, then $A(T) > 1$ and increases with each larger timescale. If events are random or evenly distributed, then $A(T) \approx 1$ across timescales (AF analysis does not distinguish between random and periodic events). As mentioned above, AF functions were divided into the lower and upper halves of timescales (timescales 1–6 and 7–11, respectively). The shorter timescales roughly correspond to variability in the timing of smaller units of speech, such as phonemes, syllables, and words, whereas the longer timescales roughly correspond to variability in the larger units of speech, such as phrases, sentences, and turns.

Lexical matching and JSD analysis

Lexical matching in the probability distributions over lemma usage was measured by quantifying overlap in the probability distributions of lemma frequencies used by the two speakers in each dyad. Words were coded in terms of their underlying lemmas using English and Spanish lemma dictionaries that replaced inflected words with their lemma roots (derived from http://www.corpora.heliohost.org). If a word was not in the dictionary, its originally transcribed form was preserved. For participants assigned to speak Spanish in the Mixed condition, one Spanish–English bilingual researcher listened to each conversation and then translated the individual Spanish words into their closest probable English counterparts. A second Spanish–English bilingual researcher reviewed these translations and resolved any discrepancies in the translations with the other researcher.

For each lemma spoken by each person in each conversation, its probability of occurring was computed as its token frequency divided by the total number of lemma tokens spoken by that person in that conversation. To illustrate some example lemmas and their frequencies, Table 2 shows the 20 most frequent lemmas for one randomly selected example dyad in the English language condition, and the same dyad in the Spanish language condition. The number of unique English words in either the English or Mixed condition was significantly higher ($M = 136.63$, $SD = 32.37$) than Spanish words in either the Spanish or Mixed condition ($M = 116.30$, $SD = 27.76$), $t(162.23) = 4.37$, $p < .001$.

The overlap between each participant's lemma distribution was quantified using JSD:

$$JSD(A, B) = \frac{1}{2}KL(A \parallel AB) + \frac{1}{2}KL(B \parallel AB)$$

$$KL(P \parallel Q) = -\sum\nolimits_i P_i \ln \frac{P_i}{Q_i}$$

JSD is the mean Kullback-Leibler divergence for each participant's probability distribution, A and B, relative to their combined probability distribution AB. JSD = 1 means totally non-overlapping frequency distributions, and JSD = 0 means identical distributions. JSD values were compared against a baseline to determine if the lemma distributions for a given dyad overlapped more than expected by chance. Using the previously mentioned translated data created by one of our bilingual researchers, in which Spanish words were translated to English words, a surrogate JSD value was determined for each participant in each conversation by pairing his or her lemma frequency distribution with that of all other participants in a different dyad, but in the same language and topic condition. Surrogate JSD values were averaged and compared per dyad.

Results

AF analyses

Mean AF functions for each of the three language conditions are shown in Figure 2. The nearly straight lines indicate roughly self-similar nested clustering across timescales. The bend in the functions suggests that clustering was more nested in the shorter timescales, as reflected in a steeper slope to the curve on the left side compared with the right. Visual inspection of the three mean AF functions suggests little effect of conversation type.

Before measuring complexity matching, we first tested whether slopes of AF functions differed by language condition or timescale. We ran a two-way repeated measures ANOVA with

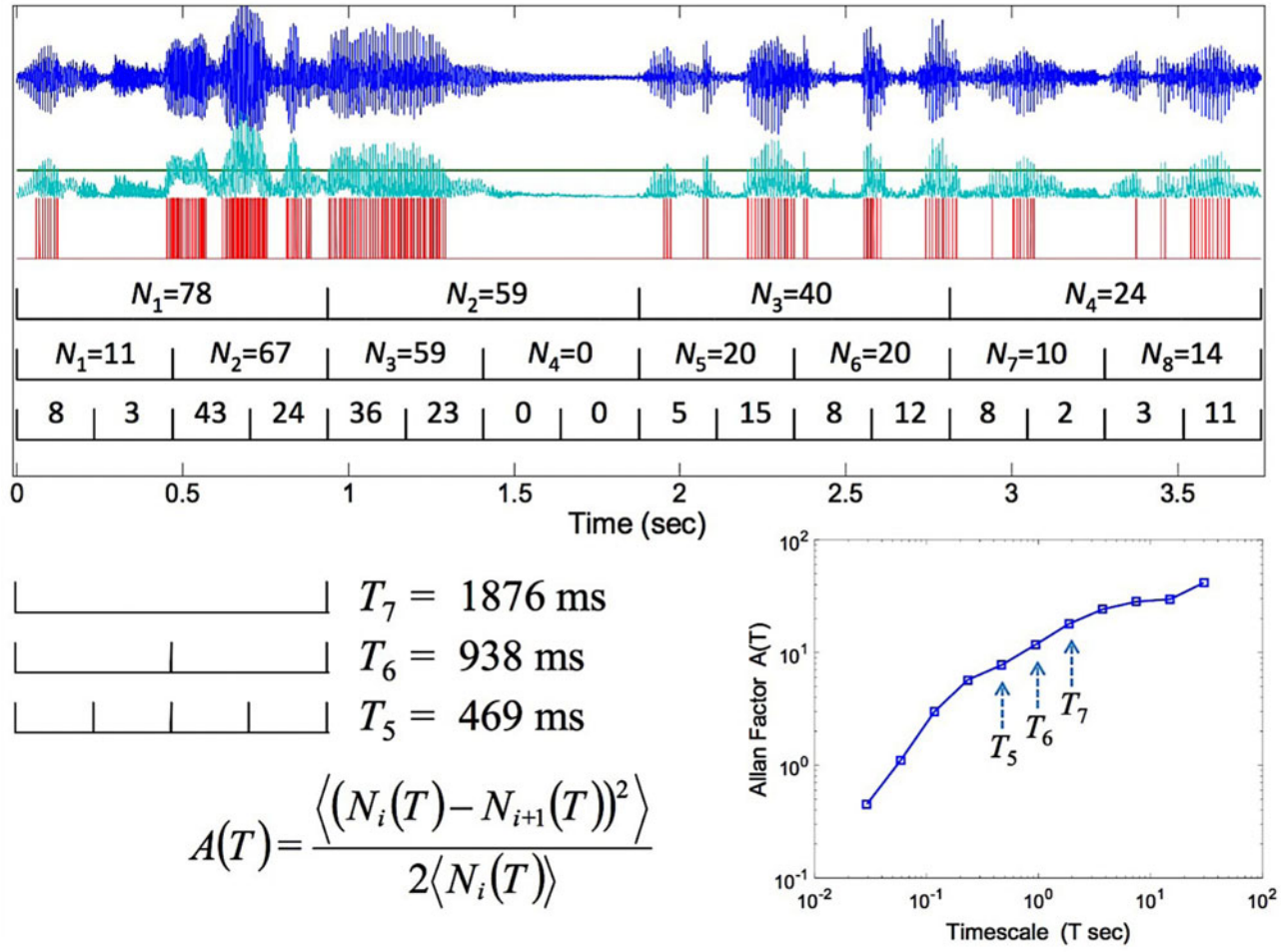

Fig. 1. Illustration of acoustic event analysis and AF analysis from Kello et al. (2017). The waveform is at the top and shows a 3.5 second waveform segment. Below the waveform is the Hilbert envelope, followed by the peak event series. Event counts *N* are shown inside brackets representing segments for three different sizes *T* (where timescale = 2*T*). Event counts *N* are set based on a threshold giving an average of one peak per 200 samples. Also shown is the AF equation and log-log plot generated from the equation, showing the amount of nested clustering at 11 different timescales.

language (English, Spanish, Mixed) and timescale (short, long) as fixed factors, the slope of the AF function in log-log coordinates as the dependent variable, and individual participants as the random factor. Slopes did not differ as a function of language ($F(2,324) = 0.41$, $p > .6$, $MSE = 0.01$), but they were steeper in the shorter timescales ($F(1,324) = 122.97$, $p < .001$, $MSE = 4.07$). There was no interaction between language and timescale ($F(2,324) = 0.18$, $p > .8$, $MSE = 0.01$).

To test for effects of order and conversation type, a three-way repeated measures ANOVA was conducted with trial number, conversational topic, and timescale as the independent variables, slope as the dependent variable, and dyad as the random factor. No main effect was found for trial, $F(1,312) = 1.52$, $p > .2$, $MSE = 0.05$, or topic, $F(2,312) = 0.05$, $p > .9$, $MSE = .002$. Additionally, no interaction was found between the three independent variables, suggesting that the effect of timescale did not influence the effects of trial or topic, $F(2,312) = 0.64$, $p > .5$, $MSE = 0.02$.

Correlations of AF slopes

Complexity matching in AF functions was analyzed using a linear mixed effects regression, which predicted one speaker's AF slope based on their partner's. In addition, we included the fixed effects of conversation order (1–3), timescale (short or long), language condition (English, Spanish, Mixed), and dyad language experience, where the latter was a binary variable indicating whether or not both participants reported Spanish as their primary, native language. Dyad was set as the random effect with a random intercept and random slope.

A reliable effect of overall complexity matching was found, B = 0.87, $t(52.8) = 18.35$, $p < .001$. The interaction between timescale and slope was added to the model to test for differences between the short and long timescales. Matching was reliable in the longer timescales (B = 0.54, $t(87.4) = 6.77$, $p < .001$) but not the shorter timescales (B = 0.18, $t(136.5) = 1.05$, $p = .3$), although the interaction with timescale was only marginally reliable (B = 0.36, $t(126.7) = 1.87$, $p = .06$; see Figure 3A). Complexity matching did not vary reliably as a function of language condition (all $p > .2$; see Figure 3B), conversation order (all $p > .3$), or conversation topic (all $p > .5$).

The results so far indicate that complexity matching does not require participants to speak the same language nor the same words or sentences. Therefore, in theory, complexity matching should be applicable to the same person compared with himself or herself speaking in two different conversations. The idea is that complexity matching may measure the character and style

Table 2. Twenty most frequent lemmas used by one dyad in two example conversations, one spoken in English and the other in Spanish. Lemmas spoken by both speakers are bolded.

English				Spanish			
Participant A		Participant B		Participant A		Participant B	
Lemma	Frequency	Lemma	Frequency	Lemma	Frequency	Lemma	Frequency
I	**29**	**be**	**30**	**que**	**18**	**no**	**12**
be	**18**	**I**	**26**	**y**	**16**	gustar	9
it	**18**	**not**	**20**	**de**	**12**	**pero**	**9**
like	**17**	**that**	**16**	sí	11	**que**	**8**
the	**16**	**it**	**16**	**escuchar**	**10**	**me**	**8**
movie	**14**	**the**	**13**	lo	9	en	7
yes	**12**	**have**	**13**	comer	9	mucho	6
of	9	a	12	el	8	entender	6
one	**9**	and	11	yo	8	**de**	**5**
that	**8**	**movie**	**11**	a	8	música	5
not	**8**	**like**	**11**	**no**	**8**	**y**	**5**
have	**7**	really	9	porque	7	su	5
do	**6**	so	9	**pero**	**7**	**escuchar**	**5**
know	6	to	9	ese	7	**ser**	**4**
you	**5**	**one**	**9**	ir	7	o	4
because	5	**but**	**8**	**me**	**6**	se	4
but	**5**	**yes**	**8**	**ser**	**6**	tener	4
they	5	**do**	**7**	estar	6	también	3
many	5	**you**	**7**	todo	5	bien	3
watch	5	good	6	tierra	5	he	3

of a person's speech, regardless of language or conversation. We consider this possibility because AF variance removes information about specific clusters of peak events, and instead gauges their variance in cluster sizes and durations. Variance in clustering may be comparable in a person's speech across different languages, even though different speech units are produced.

We tested within-speaker matching by running a linear mixed effects model to correlate each participant's AF slopes from the English and Spanish conditions with his or her AF slopes in the Mixed condition. Half of the participants spoke English in the Mixed condition, and half spoke Spanish. Therefore, each participant provided one data point speaking the same language (English and Spanish were merged for this analysis), and one data point speaking different languages. We found that AF slopes correlated within individual speakers across different conversations ($B = 0.77$, $t(37.96) = 15.63$, $p < .001$), but there was no reliable interaction with language condition ($B = -0.04$, $t(177.32) = -0.56$, $p > .05$) or timescale ($B = 0.09$, $t(184.26) = 0.57$, $p > .05$). Taken together, these results indicate that speakers exhibit patterns of nested clustering across all measured timescales that are consistent with themselves across languages and conversations.

Next, we tested whether complexity matching varied as a function of language experience. Speakers were variable in how they used the ratings scale, and ratings were mostly subjective. Instead, we chose to focus on a simple binary categorization: if both members of a dyad listed Spanish as a native language and English as a secondary language, the dyad was categorized as Spanish primary (13 dyads), and otherwise English primary (15 dyads). For the Spanish and Mixed conditions, the degree of complexity matching was not reliably affected by language experience, $B = 0.10$, $t(17.09) = 0.65$, $p > .5$. Therefore, it appears that language fluency did not vary enough in our sample of participants to affect complexity matching. Nearly all of our participants had similar language backgrounds, in that they were native Californians from families with Mexican heritage who speak a Californian dialect of Spanish and used it on a regular basis.

JSD analyses

We tested for matching in lemma usage using a three-way mixed design ANOVA, with language condition (English, Spanish, or Mixed) and JSD type (original or surrogate) as independent repeated measures factors, language experience as an independent between-subjects factor, JSD value as the dependent variable, and dyad as the random effect. A significant main effect of JSD type was found, $F(1,150) = 11.76$, $p < .01$, $MSE = 0.02$, and Figure 4 shows that original JSD values were less divergent than surrogates, indicating an overall effect of lexical matching. This effect cannot be attributed to using words that are common to a given topic of conversation because the JSD surrogates were drawn from the

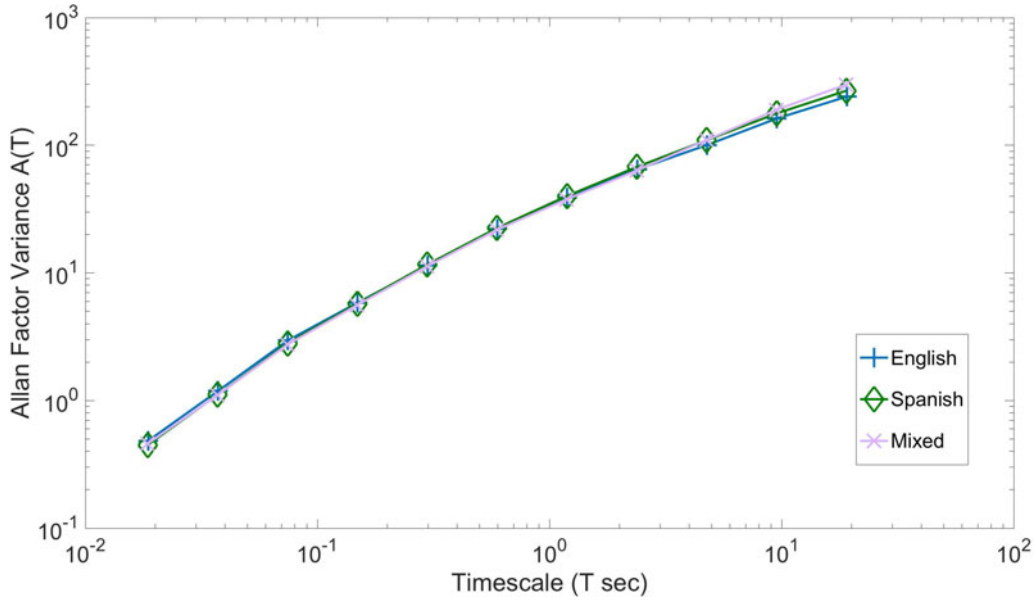

Fig. 2. Averaged Allan Factor functions showing the mean amount of nested clustering (i.e., HTS) at each timescale for each language condition.

same conversational topic as their corresponding originals. There was also a main effect of language condition, $F(2,150) = 81.42$, $p < .001$, $MSE = 0.11$, reflecting that JSD values were most divergent, or least convergent, in the Mixed condition, followed by Spanish and then English. This main effect may be due to differences in the lemma dictionaries used, and/or unavoidable issues with translation. Importantly, the difference between originals and surrogates were not reliably different for Mixed compared to pure language conditions.

To test whether lexical matching (i.e., the difference between original and surrogate JSD values) varied as a function of language condition, we again used a three-way mixed design ANOVA, with language condition (English, Spanish, or Mixed) and JSD type (original or surrogate) as independent repeated measures factors, language experience as an independent between-subjects factor, JSD value as the dependent variable, and dyad as the random effect. There was no reliable interaction between JSD type and language condition, $F(2,150) = 0.38$, $p > .6$, $MSE < 0.001$.

We also tested for effects of conversation order and topic on lexical matching using two additional three-way ANOVA models. To test for order effects, trial number, language experience, and JSD type were set as independent variables, score as the dependent variable, and dyad as the random factor. The same ANOVA was tested for conversational topic, where topic simply replaced trial number in the independent variables. Neither trial, $F(1,156) = 0.05$, $p > .8$, nor topic, $F(2,150) = 0.01$, $p > .9$, interacted with JSD type. In summary, like complexity matching, lexical matching appears to be robust to both intra- and inter-language interactions, and unaffected by variations in language experience in our participant pool.

Finally, we tested whether JSD score varied as a function of language experience and JSD type (original participant or surrogate). Again, we focused on a simple binary categorization: if both members of a dyad listed Spanish as a native language and English as a secondary language, the dyad was categorized as Spanish primary (13 dyads), otherwise English primary (15 dyads). A linear mixed effects regression tested the interaction between language experience (Spanish native or not Spanish native) and JSD type (original or surrogate). No interaction was found between these three conditions, $F(1, 100) = 0.56$, $p > .4$, nor was there a main effect of language experience, $F(1,100) = 1.82$, $p > .1$. We therefore found no evidence of an effect of language experience, consistent with the lack of effect for complexity matching (see above), which may have been due to homogeneity of language fluency in our population sample.

Relation between complexity matching and lexical matching

Analyses thus far indicate that both complexity matching and lexical matching occur in inter-language Spanish–English conversations (i.e., the Mixed condition), with no reliable difference in the magnitude of matching compared with purely English or purely Spanish conversations. For our last analyses, we examined whether complexity and lexical matching have a common basis by correlating their magnitudes. JSD difference values (original minus surrogate) are a direct measure of convergence in word usage for each given conversation, but so far we have measured complexity matching at the aggregate level in terms of AF slope correlations.

To provide a measure of complexity matching per conversation, we computed the absolute differences of AF slopes in the longer timescales produced by each dyad, which is inversely related to JSD difference scores. We then ran a linear mixed effects model with complexity matching as the dependent measure, and the negative of the JSD difference scores (to undo the inverse relationship) as the predictor, with language condition as a factor, and dyad as the random effect with a random intercept and random slope. Lexical matching was found to predict complexity matching, and vice versa, $B = .74$, $t(82.0) = 3.71$, $p < .001$.

We next tested if the relationship between complexity and lexical matching was mediated by language condition. A linear mixed effects regression was used to predict complexity matching based on the fixed effects of lexical matching and the interaction between language condition and lexical matching. Dyad was set as the random effect with a random intercept and random slope. A marginal interaction was found for the Spanish and English

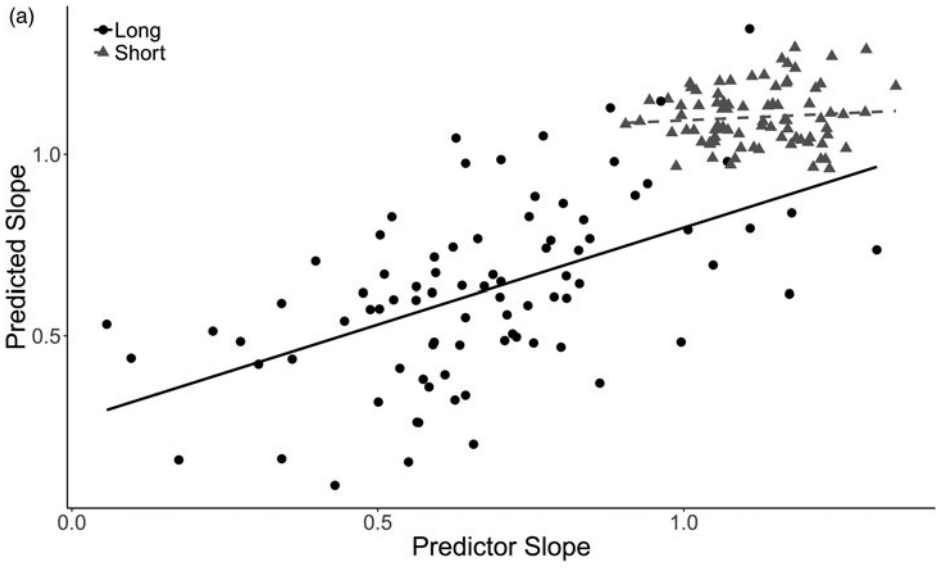

Fig. 3A. Predictor Allan Factor slopes plotted against predicted Allan Factor slopes as a function of timescale, separated for short and long timescales.

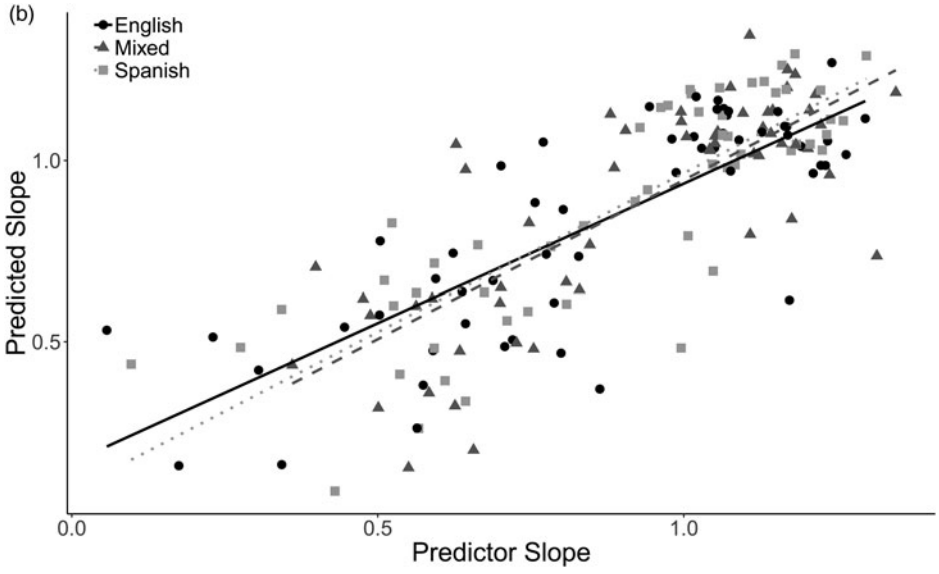

Fig. 3B. Predictor Allan Factor slopes plotted against predicted Allan Factor slopes as a function of language condition.

conditions only, B = 0.40, $t(78) = 2.06$, $p = .04$. Upon further investigation, we found that the correlation between lexical matching and complexity matching was slightly stronger for Spanish (B = −1.36) compared with English (B = −0.95). The reason for this marginal effect is unclear and its unexpectedness warrants further investigation.

The observed relationship between complexity and lexical matching does not appear to be directly causal because word durations are mostly shorter than the second+ timescales of complexity matching, and because surface forms of words are not directly matched in the Mixed condition. Therefore, overlap in the sounds of words was not the cause of complexity matching, or vice versa. Instead, it appears that convergence has an underlying basis that gives rise to both complexity and lexical matching.

Discussion

In the present study, we examined two measures of convergence in conversations spoken in English and Spanish. We measured complexity matching and lexical matching when conversational partners spoke the same language, either English or Spanish, and compared these conditions to an inter-language Mixed condition during which one partner spoke English and the other Spanish. The main result was that both types of convergence occurred in all three language conditions, and both complexity matching and lexical matching were no less prevalent in inter-language conversations. Neither type of matching was modulated by the order or topic of conversation, which taken together demonstrates the robust and general nature of convergence in conversation.

Complexity matching was reliable only in the longer timescales, in the range of hundreds of milliseconds to tens of seconds, which is consistent with prior studies indicating that prosodic and discourse processes may be more variable and therefore malleable to convergence. Complexity matching across speakers was not reliable in the shorter timescales, suggesting either that AF slopes are too coarse to detect any small-scale effects of cross-speaker convergence (e.g., in VOTs; see Balukas & Koops, 2015), or

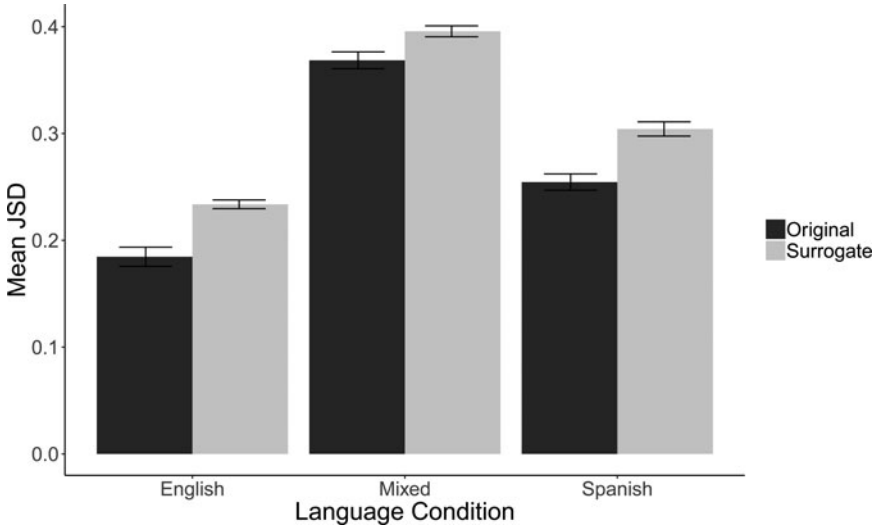

Fig. 4. Mean JSD values (with standard error bars) for original versus surrogate pairings, plotted as a function of language condition.

HTS in the shorter timescales is relatively automatized over the course of language learning. However, complexity matching was reliable WITHIN speakers when comparing a speaker with himself or herself across different conversations and languages. This intra-person convergence illustrates how complexity matching captures HTS as reflective of speech "style" rather than the particular words and other linguistic units, because we observed within-person convergence across DIFFERENT conversations, even in different languages. Finally, complexity and lexical matching were reliably correlated with each other, suggesting that they arise from common underlying processes of convergence.

The observed equivalence of speech convergence within and between languages is consistent with the prevalence of lingua receptiva across cultures and generations (ten Thije, 2013). Human language behaviors and processes appear to generalize over languages commonly and readily, despite myriad differences in linguistic units and structures that require distinct efforts devoted to learning each language. Many bilingual speakers naturally interact using both languages interchangeably, e.g., when there are asymmetries in receptive or productive speech competencies between conversational partners that lead speakers to communicate and coordinate using different languages. It is likely that lingua receptive and similar language experiences are common in the population of Californian Spanish–English bilinguals from which we sampled.

Theories of convergence generally explain the phenomenon in terms of shared activation of representations, or language processes following aligned trajectories, as in the interactive alignment theory (Garrod & Pickering, 2009; Pickering & Garrod, 2004; Trofimovich, 2016). This hypothesis appears to predict less matching across languages because at least some representations may not be directly aligned. For instance, prosody and discourse processes may be common across languages, and therefore complexity matching may be unaffected by the language spoken. To illustrate, bilingual speakers have voice qualities and styles that are recognizable across the languages spoken, and correlations in AF slopes appear to reflect a speaker's tendency to "bend" their voice towards their partner. By contrast, different languages usually have different wordform lexicons with orthographic and phonological representations that are categorical (symbolic) and cannot bend towards one another in order to align lexical representations. Therefore, our results indicate either that convergence does not rely on direct alignment – meaning the use of the same words – or language processes and representations are learned to be shared and aligned across languages for proficient bilingual speakers (Guo & Peng, 2006; Kantola & van Gompel, 2011; Kroll, Bobb & Hoshino, 2014; Marian & Spivey, 2003).

While it may be difficult to align lexical and other representations across languages, it is possible for language use to converge in the SHAPES of probability distributions over word usage and other levels of representation. Specifically, if we assume that speakers generally produce Zipfian frequency distributions (Zipf, 1935), as found in many prior studies across different languages (e.g., Peterson, Tenenbaum, Havlin, Stanley & Perc, 2012), then we may expect the shape of the power law (its exponent) produced by one speaker to bend towards that of their conversational partner, and vice versa. We could not test this hypothesis in the present study because the conversations were too short, with too few words produced per speaker, to estimate power law exponents.

In future studies, it would be interesting to measure lexical matching using Zipf's law in corpora or experiments that contain longer language interactions. It would also be interesting to test both lexical and complexity matching in bilingual interactions between pairs of languages that vary in their phonological, grammatical, and lexical similarity. Other sources of variability to be examined are the language fluencies and backgrounds of speakers. We did not find any effects of language experience on complexity or lexical matching, but our participants were drawn from a fairly homogenous population of Spanish-speaking Californians with family roots primarily in Mexico. Future studies could sample from a wider range of language fluencies, and greater variation in lingua receptiva experience, to test for effects of language experience on measures of matching.

Finally, we found no effect of conversation order on complexity matching or lexical matching, which suggests that convergence developed relatively quickly within the first five-minute conversation (see Brennan & Clark, 1996; Potter & Lombardi, 1998). However, our measure of complexity matching is ill-suited to measuring time course effects because Allan Factor analysis requires approximately four to five minutes of audio in order to accurately measure hierarchical temporal structure. Likewise, JSD requires an entire word distribution. With longer conversations, one could slide a moving window over the acoustic signal

or the time series of words spoken to measure the time course of complexity matching and lexical matching using measures like ours.

In conclusion, the present study explored convergence in monolingual and bilingual conversations using two new measures of matching. We provided evidence for complexity matching and lexical matching as general, robust phenomena in both monolingual and bilingual conversations. Together, these measures of speech convergence appear to reflect basic principles of social interaction and shared processes of monolingual and bilingual language interaction.

Supplementary Material. For supplementary material accompanying this paper, visit https://doi.org/10.1017/S1366728919000774

Acknowledgements. We thank Drew Abney and Rick Dale for providing input and assistance during the early stages of this project, and we thank Gilbert Sepulveda and Alexis Luna for their assistance in collecting data and transcribing the recordings. We also thank Kathleen Coburn for her input and statistical assistance. Finally, we would like to thank our reviewers for their detailed feedback that greatly helped to improve the paper. The research was partly supported by NSF award 1529127.

Declarations of interest. None. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Informed consent was obtained for experimentation with human subjects.

References

Abney DH, Kello CT and Warlaumont AS (2015) Production and Convergence of Multiscale Clustering in Speech. *Ecological Psychology* **27**, 222–235.

Abney DH, Paxton A, Dale R and Kello CT (2014) Complexity matching in dyadic conversation. *Journal of Experimental Psychology* **143**, 2304–2315.

Abney DH, Warlaumont AS, Oller DK, Wallot S and Kello CT (2017) Multiple Coordination Patterns in Infant and Adult Vocalizations. *Infancy: The Official Journal of the International Society on Infant Studies* **22**, 514–539. https://doi.org/10.1111/infa.12165

Anderson A, Garrod SC and Sanford AJ (1983) The accessibility of pronominal antecedents as a function of episode shifts in narrative text. *The Quarterly Journal of Experimental Psychology Section A* **35**, 427–440. https://doi.org/10.1080/14640748308402480

Bahtina-Jantsikene D and Backus A (2016) Limited common ground, unlimited communicative success. *Philologia Estonica Tallinnensis*, (1), 17–36.

Bahtina D, ten Thije JD and Wijnen F (2013) Combining cognitive and interactive approaches to lingua receptiva. *International Journal of Multilingualism* **10**, 159–180. https://doi.org/10.1080/14790718.2013.789521

Balukas C and Koops C (2015) Spanish-English bilingual voice onset time in spontaneous code-switching. *International Journal of Bilingualism* **19**, 423–443. https://doi.org/10.1177/1367006913516035

Bialystok E, Craik FIM and Luk G (2012) Bilingualism: consequences for mind and brain. *Trends in Cognitive Sciences* **16**, 240–250. https://doi.org/10.1016/j.tics.2012.03.001

Blumenfeld HK and Marian V (2007) Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking. *Language and Cognitive Processes* **22**, 633–660. https://doi.org/10.1080/01690960601000746

Bock JK (1986) Syntactic persistence in language production. *Cognitive Psychology* **18**, 355–387. https://doi.org/10.1016/0010-0285(86)90004-6

Bortfeld H and Brennan SE (1997) Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Processes* **23**, 119–147. https://doi.org/10.1080/01638537709544986

Branigan HP, Pickering MJ, Liversedge SP, Stewart AJ and Urbach TP (1995) Syntactic priming: Investigating the mental representation of language. *Journal of Psycholinguistic Research* **24**, 489–506.

Brennan SE and Clark HH (1996) Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **22**, 1482–1493.

Brennan SE, Kuhlen AK and Charoy J (2018) Discourse and dialogue. In Wixted JT and Thompson-Schill SL (eds), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, Language and Thought* (Vol. 3). Wiley.

Brennan SE and Hanna JE (2009) Partner-Specific Adaptation in Dialog. *Topics in Cognitive Science* **1**, 274–291. https://doi.org/10.1111/j.1756-8765.2009.01019.x

Buchsbaum BR, Gregory H and Colin H (2001) Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science* **25**, 663–678. https://doi.org/10.1207/s15516709cog2505_2

Clark HH and Brennan SE (1991) Grounding in communication. In Resnick L, LB, John M, Teasley S and D (eds), *Perspectives on Socially Shared Cognition*. American Psychological Association.

Costa A, Pickering M and Sorace A (2008) Alignment in second language dialogue. *Language and Cognitive Processes* **23**, 528–556.

De Looze C, Oertel C, Rauzy S and Campbell N (2011) Measuring dynamics of mimicry by means of prosodic cues in conversational speech. In *International Conference on Phonetic Sciences* (pp. 1294–1297).

De Looze C, Scherer S, Vaughan B and Campbell N (2014) Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Communication*. https://doi.org/10.1016/j.specom.2013.10.002

Falk S and Kello CT (2017) Hierarchical organization in the temporal structure of infant-direct speech and song. *Cognition* **163**, 80–86. https://doi.org/10.1016/j.cognition.2017.02.017

Fleischer Z, Pickering MJ and McLean JF (2012) Shared information structure: Evidence from cross-linguistic priming. *Bilingualism: Language and Cognition* **15**, 568–579.

Fricke M and Kootstra GJ (2016) Primed codeswitching in spontaneous bilingual dialogue. *Journal of Memory and Language* **91**, 181–201. https://doi.org/10.1016/j.jml.2016.04.003

Garrod S and Anderson A (1987) Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition* **27**, 181–218.

Garrod S and Pickering MJ (2004) Why is conversation so easy? *Trends in Cognitive Sciences* **8**, 8–11. https://doi.org///dx.doi.org/10.1016/j.tics.2003.10.016

Garrod S and Pickering MJ (2009) Joint action, interactive alignment, and dialog. *Topics in Cognitive Science* **1**, 292–304. https://doi.org/10.1111/j.1756-8765.2009.01020.x

Gries STH and Koostra GJ (2017) Structural priming within and across languages: A corpus-based perspective. *Bilingualism: Language and Cognition* **20**, 235–250. https://doi.org/10.1017/S1366728916001085

Grosjean F (2010) *Bilingual: life and reality*. Harvard University Press.

Guo T and Peng D (2006) Event-related potential evidence for parallel activation of two languages in bilingual speech production. *NeuroReport* **17**, 1757–1760.

Hardy SM, Messenger K and Maylor EA (2017) Aging and syntactic representations: Evidence of preserved syntactic priming and lexical boost. *Psychology and Aging* **32**, 588–596. https://doi.org/10.1037/pag0000180

Hartsuiker RJ, Pickering MJ and Veltkamp E (2004) Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science* **15**, 409–414.

Healey PG, Purver M and Howes C (2014) Divergence in dialogue. *Public Library of Science: One* **9**. https://doi.org/10.1371/journal.pone.0098598

Kantola L and van Gompel RPG (2011) Between- and within-language priming is the same: Evidence for shared bilingual syntactic representations. *Memory & Cognition* **39**, 276–290.

Kello CT, Dalla Bella S, Médé B and Balasubramaniam R (2017) Hierarchical temporal structure in music, speech and animal vocalizations: jazz is like a conversation, humpbacks sing like hermit thrushes. *Journal of The Royal Society Interface* **14**. https://doi.org/10.1098/rsif.2017.0231

Kim M, Horton WS and Bradlow AR (2011) Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology* **2**, 125–156.

Kroll JF, Bobb SC and Hoshino N (2014) Two Languages in Mind: Bilingualism as a Tool to Investigate Language, Cognition, and the Brain. *Current Directions in Psychological Science* **23**, 159–163. https://doi.org/10.1177/0963721414528511

Kroll JF, Dussias PE, Bogulski CA and Kroff JRV (2012) Juggling two languages in one mind: What bilinguals tell us about language processing and its consequences for cognition. *Psychology of Learning and Motivation* **56**, 229–262. https://doi.org/10.1016/B978-0-12-394393-4.00007-8

Levitan R, Benus S, Gravano A and Hirschberg J (2015) Entrainment and Turn-Taking in Human-Human Dialogue. In *AAAI Spring Symposium on Turn-Taking and Coordination in Human-Machine Interaction*.

Lowen SB and Teich MC (1996) The periodogram and Allan variance reveal fractal exponents greater than unity in auditory-nerve spike trains. *The Journal of the Acoustical Society of America* **99**, 3585–3591.

Luque J, Luque B and Lacasa L (2015) Scaling and universality in the human voice. *Journal of The Royal Society Interface* **12**. https://doi.org/10.1098/rsif.2014.1344

Manson JH, Bryant GA, Gervais MM and Kline MA (2013) Convergence of speech rate in conversation predicts cooperation. *Evolution and Human Behavior* **34**, 419–426. https://doi.org/10.1016/j.evolhumbehav.2013.08.001

Marian V and Spivey M (2003) Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition* **6**, 97–115. https://doi.org/DOI: 10.1017/S1366728903001068

Marmelat V and Delignières D (2012) Strong anticipation: Complexity matching in interpersonal coordination. *Experimental Brain Research* **222**, 137–148.

Nenkova A, Gravano A and Hirschberg J (2008) High frequency word entrainment in spoken dialogue. In *Proceedings of the 46th annual meeting of the association for computational linguistics on human language technologies: Short papers* (pp. 169–172). Association for Computational Linguistics.

Ni Eochaidh C (2010) *The role of conceptual and word form representations in lexical alignment: Evidence from bilingual dialogue*. The University of Edinburgh.

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

Nielsen K (2011) Specificity and abstractness of VOT imitation. *Journal of Phonetics* **39**, 132–142. https://doi.org/10.1016/j.wocn.2010.12.007

Pardo J (2013) Measuring phonetic convergence in speech production. *Frontiers in Psychology* **4**.

Pardo JS (2006) On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America* **119**, 2382–2393.

Pardo JS, Gibbons R, Suppes A and Krauss RM (2012) Phonetic convergence in college roommates. *Journal of Phonetics* **40**, 190–197. https://doi.org/10.1016/j.wocn.2011.10.001

Pardo J, Urmanache A, Gash H, Wiener J, Mason N, Wilman S, Francis K and Decker A (2018) The Montclair map task: Balance, efficacy, and efficiency in conversational interaction. *Language and Speech*, 1–21. https://doi.org/10.1177/0023830918775435

Park SM and Sarkar M (2007) Parents' attitudes toward heritage language maintenance for their children and their efforts to help their children maintain the heritage language: A case study of Korean-Canadian immigrants. *Language, Culture and Curriculum* **20**, 223–235. https://doi.org/10.2167/lcc337.0

Perani D, Abutalebi J, Paulesu E, Brambati S, Scifo P, Cappa SF and Fazio F (2003) The role of age of acquisition and language usage in early, high-proficient bilinguals: An fMRI study during verbal fluency. *Human Brain Mapping* **19**, 170–182. https://doi.org/10.1002/hbm.10110

Peterson AM, Tenenbaum JN, Havlin S, Stanley HE and Perc M (2012) Languages cool as they expand: Allometric scaling and the decreasing need for new words. *Scientific Reports* **2**, 943.

Pickering MJ and Garrod S (2004) Toward a mechanistic psychology of dialogue. *Behavioral & Brain Sciences* **27**, 169–226.

Potter MC and Lombardi L (1998) Syntactic priming in immediate recall of sentences. *Journal of Memory and Language* **38**, 265–282.

Romaine S (2012) *The bilingual and multilingual community*. (T. K. Bhatia & W. C. Ritchie, Eds.). Malden, MA: Blackwell Publishing. https://doi.org/10.1002/9781118332382.ch18

Sancier ML and Fowler CA (1997) Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* **25**, 421–436. https://doi.org/10.1006/jpho.1997.0051

Schober MF and Clark HH (1989) Understanding by addressees and overhearers. *Cognitive Psychology* **21**, 211–232. https://doi.org/10.1016/0010-0285(89)90008-X

ten Thije J D, Gooskens C, Daems F, Cornips L and Smits M (2017) Lingua receptiva: Position paper on the European commission's skills agenda. *European Journal of Applied Linguistics* **5**, 141–146.

ten Thije JD (2013) Lingua receptiva (LaRa). *International Journal of Multilingualism* **10**, 137–139. https://doi.org/10.1080/14790718.2013.789519

Tian X and Poeppel D (2012) Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. *Frontiers in Human Neuroscience* **6**.

Tobin SJ, Nam H and Fowler CA (2017) Phonetic drift in Spanish-English bilinguals: Experiment and a self-organizing model. *Journal of Phonetics* **65**, 45–59. https://doi.org/10.1016/j.wocn.2017.05.006

Toribio AJ (2004) Convergence as an optimization strategy in bilingual speech: Evidence from code-switching. *Bilingualism: Language and Cognition* **7**, 165–173. https://doi.org/10.1017/S1366728904001476

Trofimovich P (2016) Interactive alignment: A teaching-friendly view of second language pronunciation learning. *Language Teaching* **49**, 411–422. https://doi.org/10.1017/S0261444813000360

van Baaren RB, Holland RW, Steenaert B and van Knippenberg A (2003) Mimicry for money: Behavioral consequences of imitation. *Journal of Experimental Social Psychology* **39**, 393–398. https://doi.org/10.1016/S0022-1031(03)00014-3

West BJ, Geneston EL and Grigolini P (2008) Maximizing information exchange between complex networks. *Physics Reports* **468**, 1–99.

Xia Z, Levitan R and Hirschberg J (2014) Prosodic entrainment in Mandarin and English: A cross-linguistic comparison. In *Proceedings of the 7th International Conference on Speech Prosody* (pp. 65–69).

Zipf GK (1935) *The psycho-biology of language*. Oxford, England: Houghton, Mifflin.

Appendix A: Pre-Experiment Language History Questionnaire

1. SONA ID:
2. Gender:
3. Age:
4. Do you have any visual and/or hearing problems? If yes, what are they?
5. What is your native country/ies?
6. What is your native langue(s)?
7. What language is spoken in your household?
8. At what age(s) did you start to learn each language, and for how many years?
9. What would you consider to be your primary second language?
10. What language are you most comfortable using on a daily basis?
11. On a scale of one to ten, with ten being the highest level of confidence, please mark your proficiency in the following areas:
    a. English reading
       1 2 3 4 5 6 7 8 9 10
    b. English spelling
       1 2 3 4 5 6 7 8 9 10
    c. English writing
       1 2 3 4 5 6 7 8 9 10
    d. English speaking
       1 2 3 4 5 6 7 8 9 10
    e. English speech comprehension
       1 2 3 4 5 6 7 8 9 10

12. On a scale of one to ten, with ten being the highest level of confidence, please mark your proficiency in the following areas:
    a. Spanish reading
       1 2 3 4 5 6 7 8 9 10
    b. Spanish spelling
       1 2 3 4 5 6 7 8 9 10
    c. Spanish writing
       1 2 3 4 5 6 7 8 9 10
    d. Spanish speaking
       1 2 3 4 5 6 7 8 9 10
    e. Spanish speech comprehension
       1 2 3 4 5 6 7 8 9 10
13. Estimate, in terms of percentages, how often you use your dominant language and other languages **per week** (in all weekly activities combined, circle which range best applies):

    Dominant language: 0% 0–25% 50–75% 75–100%
    Second language: 0% 0–25% 50–75% 75–100%

Appendix B: Post-Experiment Questionnaire

1. Have you ever met your partner before today? If so, are you just acquaintances, or friends?
2. On a scale of 1 to 5, how easy was the conversation in which you both spoken English, with 5 being the easiest?
   1 2 3 4 5
3. On a scale of 1 to 5, how easy was the conversation in which you both spoke Spanish, with 5 being the easiest?
   1 2 3 4 5
4. On a scale of 1 to 5, how easy was the conversation in which you spoke two different languages, with 5 being the easiest?
   1 2 3 4 5