INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

Lexical Entrainment and Intra-Speaker Variability in Cooperative Dialogues

Alla Menshikova, Daniil Kocharov, Tatiana Kachkovskaia

Saint Petersburg State University, Saint Petersburg, Russia

menshikova.alla2016@yandex.ru, kocharov@phonetics.pu.ru, kachkovskaia@phonetics.pu.ru

Abstract

In dialogues, intra-speaker variability is often explained by the relationship between interlocutors. A person may speak differently with a friend and a stranger or depending on the interlocutor's gender or age—in all these cases we expect speech entrainment, but the degree of entrainment may vary. In this research, we measured lexical entrainment in a series of dialogues, where each one of 20 "core" speakers talked to five different interlocutors: a sibling, a close friend, an unfamiliar person of the same gender and similar age, an unfamiliar person of the other gender and similar age, and an unfamiliar person of the same gender, greater age and higher job position. We hypothesized that the degree of speech entrainment systematically varies according to the type of interlocutor, across all the "core" speakers. The following measures of entrainment were used: parts of speech statistics, verb forms statistics, language style matching, and lexical density. Our data have shown that a person speaks very similarly to his/her sibling; dialogues with a friend or a same-gender stranger of similar age show fewer similarities; the least "common language" is observed in dialogues with a stranger of the opposite gender and with a stranger of greater age and higher job position.

Index Terms: speech entrainment, intra-speaker variability, lexical entrainment, lexical density, language style matching

1. Introduction

In recent years a lot of research has been published on the question of how speakers change their voices depending on *who* they are talking to [1, 2, 3, 4, 5]. The phenomenon of such attuning is termed "speech entrainment" (as well as "accommodation", "alignment", etc.). It has been shown that the degree of speech entrainment may depend on a number of situational, social, and individual factors. In task-oriented dialogues, task success is correlated with the degree of entrainment: the more speakers entrain, the better they perform the given task [6]. It has been pointed out that speakers' conversational roles (information giver vs. information receiver) should not be ignored within the analysis [1, 7, 8, 4]. Among the social factors, gender and social distance (familiarity, social hierarchy) are reported to play a role [1, 7, 4]. There are also individual features that affect the degree of entrainment, such as "phonetic talent", attention and others [9]. Speech entrainment is manifested at all levels of linguistic analysis: syntactic [10], lexical [11], and phonetic [2]. This paper deals with lexical entrainment. But, along with lexemes, we are also analysing word forms—which makes this research related to syntax and pragmatics as well.

A number of studies deal with the choice between synonyms. They all show that interlocutors tend to use the same lexical item within the possible ones. In [12] the authors claim that the variability of lexical choice for a particular object is rather high between two different dialogues, but significantly lower within the same dialogue. That is, when two people talk about the same objects for some time, they develop a common set of terms to describe them. At the same time, the *degree* of lexical entrainment may vary from dialogue to dialogue. In [13], where the measure of entrainment is based on the relative frequency of the most frequent words in the interlocutors' speech, the authors found significant correlations between lexical entrainment and two factors: perceived naturalness of the dialogue and task success.

One can also trace speech entrainment within the *syntactic* level of analysis. In a controlled experiment, [14] showed that the syntactic structure of the interlocutor's speech systematically varies depending on the syntactic structure of the other interlocutor's speech. Another controlled experiment confirming the idea of syntactic priming is given in [15]. They showed that a complex noun phrase with a relative clause (e.g., "the square that's red") is used more often after one hears a syntactically similar noun phrase (rather than after hearing a simple noun phrase).

Various measures have been used to estimate the degree of lexical and syntactic entrainment. Since the mechanisms underlying speech entrainment are probably related to priming, one of the approaches to measure lexical and syntactic entrainment consists in calculating the probability of word repetitions within an N-word window (the first word being considered the potential prime) [16]. Similarly, one can count repetitions of syntactic rules [10]. A numeric measure of lexical similarity between the interlocutors' speech, called language style matching (LSM), was used in [11]. In this study, based on recordings of dialogues during speed dating, the authors showed that LSM values significantly predicted relationship initiation. That is, this measure seems to reflect mutual likability within the dialogue. LSM was also successfully used in [17], where the authors analysed written conversations (chatting) in English in terms of a number of linguistic, social, and cognitive categories. Among the linguistic factors were the percentage of prepositions, past tense verbs, present-tense verbs, negations, articles, and others. Interestingly, here the authors used a different way of calculating LSM: on a turn-by-turn level. That is, the values for each of the odd turns were compared with the respective values for the adjacent even turns by means of correlation analysis; thus, this variant of LSM is a measure of synchrony within the dialogue.

The phenomenon of speech entrainment is closely related to the problem of intra-speaker (within-speaker) variability. In this paper, we addressed the question of how one's speech can change in a set of dialogues with different conversational partners. We assume that in these dialogues the social distance between the interlocutors gradually grows. As a result, the material was analysed from two perspectives—both speech entrainment and intra-speaker variability—by applying the methods traditionally used in speech entrainment research.

http://dx.doi.org/10.21437/Interspeech.2021-1441

2. Material

The experiments described in this paper were conducted using the SibLing Corpus of Russian Dialogue Speech [18]. The corpus contains studio recordings of task-oriented dialogues between Russian native speakers. Subjects completed two collaborative tasks—a card matching game and a map task. The card matching game was based on searching for similarities in two decks of cards. The interlocutors described their pictures to each other in order to find matching elements on them (completion time—10 to 15 min). In the map task, the speakers were asked to guide each other through a set of schematic maps (completion time—15 to 60 min). During the recording, the subjects were separated by a non-transparent screen.

The corpus is made up of 90 dialogues produced by 10 pairs of same-gender siblings of similar age, 24–40 y.o.: 5 male pairs and 5 female pairs; four of the sibling pairs are monozygotic twins. These were the core speakers of the corpus. Each sibling participated in 5 dialogues: (1) with the other sibling, (2) a close friend of the same gender and similar age, (3) a stranger of the same gender and similar age, (4) a stranger of the other gender and similar age, (5) a stranger of the same gender but of greater age and higher job position.

For this research, we used all dialogues produced by 10 pairs of siblings. The total number of running words is 267,280. The mean number of running words for the card-matching task is 758 (STD = 339), the mean number of running words for the map task is 1914 (STD = 1147).

3. Method

3.1. Data preprocessing

In the following steps of the analysis, the two collaborative tasks—card matching game and map task—were treated separately, due to the fact that the map task has clearly pre-defined roles (information giver vs. information receiver), while the card matching game is freer in this sense and is sometimes very close to the free conversation. This difference might lead to different entrainment results as speakers' roles influence the speech of the players [1]. There is also evidence that map task and spontaneous speech differ in between-speaker priming [10].

Orthographic transcriptions were processed by means of pymorphy2, a morphological analyzer for Python [19]. The tool was applied for word lemmatization and extraction of morphological and part-of-speech tags.

3.2. Features

3.2.1. POS-dependent bag-of-words

The choice of words for the designation of objects, actions, and characteristics can serve as evidence of successful communication. Thus, we decided to use the relative frequency of lexemes to measure speech entrainment. We divided lexemes into three groups: designating objects – nouns (N); designating characteristics – adjectives, participles, and adverbs (A); designating actions – verbs and adverbial participles (V). We calculated POS-dependent bag-of-words embeddings through the following steps: (1) the full set of lexemes encountered in all transcriptions for the given task was compiled, and (2) for each interlocutor, the embedding was constructed of the values corresponding to the number of times he/she pronounced each word in the given dialogue. As the dialogues differed in duration, all the values were normalized by the total number of word tokens pronounced by the speaker in the given dialogue.

3.2.2. Verb forms

Russian is a language with complex verb morphology that includes grammatical categories of person, gender, number, mood, etc. Hence, verb forms might also be used to estimate speech entrainment. It is particularly important that verb morphological categories may be used to express various degrees of politeness and social attitudes, especially person, number, mood—e.g., the second-person plural form is usually used in conversations with unfamiliar or elder people.

For each speaker in each dialogue, we composed a feature embedding 'V.morph' representing the speaker's use of verbs in the dialogue, composed of:

- the total number of verbs and adverbial participles;
- the number of non-finite verb forms;
- the number of finite verb forms;
- the number of adverbial participles;
- the number of imperatives;
- the number of second-person singular verbs;
- the number of second-person plural verbs;
- the number of first-person plural verbs (they are sometimes used as an indirect replacement of imperatives, e.g. "Идём!" ("Let's go!"));
- the number of first-person singular verbs;
- the number of forms "давай, давайте" ("let us [do something]").

Each value was additionally normalized by the total number of word tokens pronounced by the speaker in the given dialogue.

3.2.3. Language style matching

Language style matching (LSM) score was calculated using the method suggested in [11]. We used the following list of function word categories: pronouns, conjunctions, prepositions, negations, particles, interjections.

For each speaker, we calculated the percentage of words within each category. Then, the LSM score for two speakers was calculated using the following formula:

$$LSM = \frac{1}{N}\sum_{i=0}^{N} \frac{|words_1^{cat_i} - words_2^{cat_i}|}{words_1^{cat_i} + words_2^{cat_i} + 0.0001} \quad (1)$$

In this formula, N is the total number of categories, $words_1^{cat_i}$ is the percentage of words of the current category used by the first speaker, $words_2^{cat_i}$ is the percentage of words of the current category used by the second speaker. A higher value represents a greater distance between the speakers.

3.2.4. Lexical density

In applying the feature called lexical density we followed the paper [20]. This feature was used to estimate lexical richness of the speaker using the following formula:

$$LD = \frac{lexemes_n}{tokens_n} \quad (2)$$

where $lexemes_n$ is the number of unique words pronounced by the speaker, and $tokens_n$ is the total number of word tokens pronounced by the speaker.

Table 1: *Number of speakers (out of 20) with significant Pearson's correlation ($p < 0.05$) between lexical features and the assumed social distance between the interlocutors, in card matching game (before slash) and map task (after slash), using 3 distance measures. Empty cells stand for 0/0, minuses—for 0.*

	N	Adj	V	V.morph	LSM	LD
D1			2 / –	4 / 7	5 / 4	5 / –
D2		1 / 7	2 / 3	2 / 4	3 / 1	2 / 1
D3	1 / 2	– / 1	– / 1	– / 2	2 / –	1 / 1

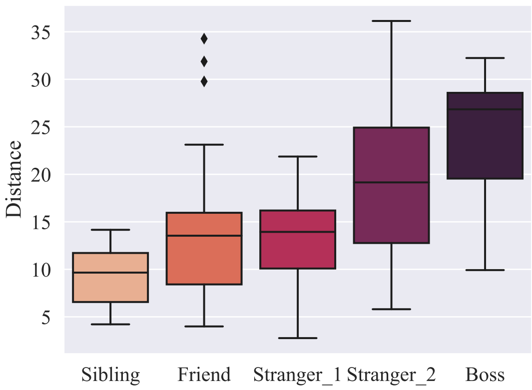

Figure 1: *Boxplots for lexical distance between speakers using D2 measure calculated for verb morphological tags in map task.*

3.3. Feature analysis

The analysis was based on comparing pairs of speakers in terms of the lexical features described above. Manhattan distance was used as the metric of the distance between feature embeddings. For lexical density and LSM, distances were calculated as absolute differences between the two values.

There are several ways to estimate intra-speaker variability and speech entrainment—depending on which speech recordings to compare. Below we describe three measures of distance between the interlocutors: D1, D2, and D3.

D1 (intra-speaker variability) reflects how a person's speech differs across all the dialogues where he/she participated. D1 is calculated as distances between the core speaker in the sibling-sibling dialogue and the same speaker (himself/herself) in all the other dialogues. This results in 4 values. We hypothesize that given a speaker, their manner of communication with various interlocutors would change. That is, the more numerous the extralinguistic differences between speakers (degree of familiarity, same vs. different gender, age, etc.), the more the speaker is supposed to change their speech due to entrainment.

D2 (speaker difference) is aimed to show the degree of difference between the speaker and his/her interlocutors. Here we intentionally disregard the changes in the speech of the core speaker and only analyse the variability in the 5 interlocutor's speech. It is calculated as distances between the core speaker in the sibling-sibling dialogue and his/her interlocutors in all the other dialogues (5 values in total). As in the previous case, we hypothesize that the distance should grow along with the social distance between the speakers.

D3 (interlocutor difference) is aimed to show the degree of mutual entrainment between interlocutors. It is calculated

Table 2: *ANOVA results representing the influence of the interlocutor factor on lexical entrainment, in card matching game (before slash) and map task (after slash), using 3 distance measures. Single asterisks correspond to $p < 0.05$, double—to $p < 0.01$; empty cells and minuses stand for no significance.*

	N	Adj	V	V.morph	LSM	LD
D1	** / *	* / *	– / **	– / **		– / **
D2	** / **	** / **	– / **	– / **	** / **	* / **
D3	** / –	– / **	– / *	– / *	– / **	

as distances between the core speaker in each dialogue and their interlocutors in the corresponding dialogues. In this case, we measure the degree of morpho-lexical "closeness" between speakers in the same dialogue. As opposed to the previous sets of distances, we do not expect to observe a gradual growth of distance values here, as the core speaker might show strong entrainment to the interlocutor despite the social distance (or even due to social distance).

3.4. Statistical analysis

We tested the hypothesis that the distance between the speaker and their interlocutors is growing from the first dialogue (with the sibling) to the last (with the person of greater age). For every distance measure, Pearson's correlation was used to test the presence of a rising trend in the line of distance values. However, since the first hypothesis could turn out to be too strong, we ran one-factor repeated-measures ANOVAs to check if there are significant differences at all between groups of distance values for different types of dialogues. Post-hoc paired Student's t-tests were run for those ANOVA test settings that were found to have a significant difference between the groups.

4. Results and discussion

Correlation analysis does not show significance in the majority of cases: the maximum number of dialogues with significant correlations is only 7 (out of 20) (Table 1). Obviously, our assumed order of social distance—from sibling to a stranger of greater age (termed "boss")—does not correlate with lexical distance in the way we expected. However, the distances may still show a systematic pattern in terms of the relationship between the interlocutors.

Indeed, ANOVA confirms the latter statement: most cells in Table 2 show significant differences, with the most filled cells for D2 distance measure. Post-hoc t-tests (Table 3) enable us to see which pairs of speakers are the source of this difference.

Sibling-sibling dialogues stand much apart from the other ones. All lexical features show significant differences in map task (with negative differences of the means — see Table 3, measures D2 and D3). That is, our data confirm that siblings speak in a similar way. The high significance of the difference between sibling-sibling dialogues and dialogues with other speakers might indicate that siblings have a "pre-defined" common language.

The core speaker's language is more similar to that of a friend or a same-gender stranger of similar age ("stranger1") compared with that of other strangers ("stranger2" and "boss"). Statistically significant differences are observed for nouns and verb morphology (see Table 3, measure D1). At the same time, the highest lexical diversity (LD) is observed in conversations with a close friend. That is, friends may also have a special

Table 3: *Statistically significant differences between dialogues with different interlocutors measured as paired t-tests across different features and 3 distance measures, in card matching games (before slash) and map task (after slash). Symbols $<$ and $>$ reflect the difference of the group means (smaller vs. greater). Single symbols correspond to $p < 0.05$, double—to $p < 0.01$; empty cells and minuses stand for no significance.*

Measure	1st group	2nd group	N	A	V	V. morph	LSM	LD
D1	friend	stranger1	$> / >$					$- / \gg$
	friend	stranger2				$- / <$		$- / \gg$
	friend	boss		$< / -$		$- / \ll$		$- / \gg$
	stranger1	stranger2		$- / <$				
	stranger1	boss	$\ll / \ll$	$< / \ll$		$- / \ll$		
	stranger2	boss		$< / -$				
D2	sibling	friend	$\ll / \ll$	$\ll / \ll$	$- / \ll$	$- / \ll$	$- / \ll$	$< / \ll$
	sibling	stranger1	$\ll / \ll$	$\ll / \ll$	$- / \ll$	$- / \ll$	$\ll / \ll$	$\ll / \ll$
	sibling	stranger2	$\ll / \ll$	$\ll / \ll$	$- / \ll$	$- / \ll$	$\ll / \ll$	$< / \ll$
	sibling	boss	$\ll / \ll$	$\ll / \ll$	$- / \ll$	$- / \ll$	$\ll / \ll$	$\ll / <$
	friend	stranger1	$> / -$					
	friend	stranger2	$> / >$					
	friend	boss		$- / <$	$- / \ll$	$- / \ll$	$\ll / -$	
	stranger1	stranger2				$- / \ll$		
	stranger1	boss		$- / <$	$- / \ll$	$- / \ll$		
	stranger2	boss	$- / <$				$< / -$	
D3	sibling	friend	$\ll / -$	$- / \ll$		$- / \ll$	$- / \ll$	
	sibling	stranger1	$< / -$	$- / \ll$		$- / <$	$- / \ll$	
	sibling	stranger2		$- / \ll$	$- / <$	$- / \ll$	$- / \ll$	
	sibling	boss	$< / -$	$- / \ll$	$- / \ll$	$- / \ll$	$- / \ll$	
	stranger1	boss			$- / <$			

common language that is richer in vocabulary.

Another conclusion that can be inferred from Table 3 concerns dialogues with a close friend compared with dialogues with a stranger of the same gender and similar age (termed "stranger1"). Two out of three distance measures show that the core speaker is closer to a stranger than to a friend; this difference is significant for nouns and lexical density. This is probably why the correlation analysis showed so few significant cases: our assumed order of interlocutors in terms of social distance is actually different.

Another result that can be seen from Table 3 concerns the difference between dialogues with a stranger of the opposite gender and similar age ("stranger2") and with a stranger of greater age and higher job position ("boss"). Very few lexical features show a statistically significant difference between these two dialogues; if we compare other dialogues with these two, we can also see that these two dialogues show a very similar pattern. That is, roughly speaking, the core speaker is as close to "stranger2" as he is to the "boss".

The actual order of social distances—from smaller to greater—can be demonstrated by Figure 1 which shows boxplots for verb morphology using the D2 measure. The friend-stranger1 difference is not significant there (see Table 3), but one can still see that the values for "stranger1" violate the assumed rising trend. Thus, non-familiarity by itself does not add more distance, but may even make the speakers closer than friends. But when we add another difference—gender or social hierarchy—the distance between the speakers rises.

It should be noted here that we did not actually measure lexical entrainment assuming that in real dialogues there is always some entrainment (which has been proved in multiple papers), but the degree of entrainment may change under various factors. However, D3 data (see Table 3, D3), which should reflect differences in speech entrainment, does not show many significant differences. Apart from the conclusion that a person speaks similarly to his/her sibling, the only significant difference is observed for nouns between the groups "stranger1" and "boss". At the same time, D1 and D2 show more significant differences and also reflect speech entrainment: D1 proves that the core speaker changes his/her speech in different dialogues, while D2 reflects changes in their interlocutors' speech.

As expected, our results differ for the two types of tasks—card matching game and map task (see Table 3, before the slash and after the slash, resp.). There are more significant differences for verb features in map tasks, while the card matching game shows more differences in nouns. This is probably because in a card matching game the players do not need many verbs to find common objects on their cards, but what they definitely need are nouns. In map tasks, describing a route definitely requires using a lot of verbs of motion.

5. Conclusions

In this paper, we have shown that a person's speech varies in dialogues with different conversational partners. In terms of several lexical measures, our data have shown that a person speaks very similarly to his/her sibling; dialogues with a friend or a same-gender stranger of similar age show fewer similarities; the least "common language" is observed in dialogues with a stranger of the opposite gender and with a stranger of greater age and higher job position. This intra-speaker variability is probably the result of variation in the degree of speech entrainment between the interlocutors depending on social distance.

6. Acknowledgements

This work is supported by the Russian Science Foundation (grant 19-78-10046).

7. References

[1] J. S. Pardo, “On phonetic convergence during conversational interaction,” *Journal of Acoustical Society of America*, vol. 119, no. 4, pp. 2382–2393, 2006.

[2] R. Levitan and J. Hirschberg, “Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions,” in *Proceedings of Interspeech*, 2011, pp. 3081–3084.

[3] M. Babel and D. Bulatov, “The role of fundamental frequency in phonetic accommodation,” *Language and Speech*, vol. 55, pp. 231–248, 2012.

[4] A. Weise, S. I. Levitan, J. Hirschberg, and R. Levitan, “Individual differences in acoustic-prosodic entrainment in spoken dialogue,” *Speech Communication*, vol. 115, pp. 78–87, 2019.

[5] A. Weise, V. Silber-Varod, A. Lerner, J. Hirschberg, and R. Levitan, ““talk to me with left, right, and angles”: Lexical entrainment in spoken hebrew dialogue,” in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 2021, pp. 292–299.

[6] U. Reichel and J. Cole, “Entrainment analysis of categorical intonation representations,” in *Proceedings of of Phonetik und Phonologie im deutschsprachigen Raum*, 2016, pp. 165–168.

[7] C. Danescu-Niculescu-Mizil, L. J. Lee, B. Pang, and J. M. Kleinberg, “Echoes of power: language effects and power differences in social interaction,” in *Proceedings of the 21st International Conference on World Wide Web*, 2012, pp. 699–708.

[8] J. Abel and M. Babel, “Cognitive load reduces perceived linguistic convergence between dyads,” *Language and Speech*, vol. 60, no. 3, pp. 479–502, 2017.

[9] N. Lewandowski and M. Jilka, “Phonetic convergence, language talent, personality and attention,” *Frontiers in Communication*, no. 4, 2019.

[10] D. Reitter, J. D. Moore, and F. Keller, “Priming of syntactic rules in task-oriented dialogue and spontaneous conversation,” in *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 2006, pp. 685–690.

[11] M. E. Ireland, R. B. Slatcher, P. W. Eastwick, L. E. Scissors, E. J. Finkel, and J. W. Pennebaker, “Language style matching predicts relationship initiation and stability,” *Psychological Science*, vol. 22, no. 1, pp. 39–44, 2011.

[12] S. E. Brennan, “Lexical entrainment in spontaneous dialog,” in *Proceedings of International Symposium on Spoken Dialogue (ISSD 96)*, 1996, p. 41–44.

[13] A. Nenkova, A. Gravano, and J. Hirschberg, “High frequency word entrainment in spoken dialogue,” in *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2008, pp. 169–172.

[14] H. Branigan, M. Pickering, and A. Cleland, “Syntactic coordination in dialogue,” *Cognition*, vol. 75, pp. b13–b25, 2000.

[15] A. A. Cleland and M. J. Pickering, “The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure,” *Journal of Memory and Language*, vol. 49, no. 2, pp. 214–230, 2003.

[16] A. Ward and D. J. Litman, “Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora,” in *SLaTE Workshop on Speech and Language Technology in Education*, 2007.

[17] K. G. Niederhoffer and J. W. Pennebaker, “Linguistic style matching in social interaction,” *Journal of Language and Social Psychology*, vol. 21, no. 4, pp. 337–360, 2002.

[18] T. Kachkovskaia, T. Chukaeva, V. Evdokimova, P. Kholiavin, N. Kriakina, D. Kocharov, A. Mamushina, A. Menshikova, and S. Zimina, “SibLing corpus of Russian dialogue speech designed for research on speech entrainment,” in *Proceedings of the 12th Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, May 2020, pp. 6556–6561.

[19] M. Korobov, “Morphological analyzer and generator for Russian and Ukrainian languages,” in *Proceedings of International Conference on Analysis of Images, Social Networks and Texts*, 2015, pp. 320–332.

[20] V. Silber-Varod, S. Malayev, and A. Lerner, “Positioning oneself in different roles: Structural and lexical measures of power relations between speakers in Map Task Corpus,” *Speech Communication*, vol. 117, pp. 1–12, 2020.