Comparison of Lexical Alignment with a Teachable Robot in Human-Robot and Human-Human-Robot Interactions

Yuya Asano[1], Diane Litman[1,2,3], Mingzhi Yu[2], Nikki Lobczowski[3], Timothy Nokes-Malach[3,4], Adriana Kovashka[1,2], and Erin Walker[1,2,3]

[1]Intelligent Systems Program, University of Pittsburgh, USA
[2]Department of Computer Science, University of Pittsburgh, USA
[3]Learning Research and Development Center, University of Pittsburgh, USA
[4]Department of Psychology, University of Pittsburgh, USA
{yua17, dlitman, miy39, ngl13, nokes, aik85, eawalker}@pitt.edu

Abstract

Speakers build rapport in the process of aligning conversational behaviors with each other. Rapport engendered with a teachable agent while instructing domain material has been shown to promote learning. Past work on lexical alignment in the field of education suffers from limitations in both the measures used to quantify alignment and the types of interactions in which alignment with agents has been studied. In this paper, we apply alignment measures based on a data-driven notion of shared expressions (possibly composed of multiple words) and compare alignment in one-on-one human-robot (H-R) interactions with the H-R portions of collaborative human-human-robot (H-H-R) interactions. We find that students in the H-R setting align with a teachable robot more than in the H-H-R setting and that the relationship between lexical alignment and rapport is more complex than what is predicted by previous theoretical and empirical work.

1 Introduction and Related Work

Alignment is the convergence of behavior among speakers and plays an important role in designing the strategies of dialogue systems because it is associated with user engagement (Campano et al., 2015) and task success (Nenkova et al., 2008; Callejas et al., 2011; Lubold et al., 2018; Kory-Westlund and Breazeal, 2019). However, few studies have looked at how this relationship differs in multi-party versus dyadic task-oriented dialogues involving humans and a dialogue agent. This gap prevents us from inferring appropriate alignment strategies for dialogue agents across different group sizes.

Teachable agents act as peers that learners teach via dialogue. These agents have been shown to facilitate learning due to the effect of learning by teaching (Leelawong and Biswas, 2008) and the rapport the agents build with learners (Gulz et al., 2011). Inspired by theories suggesting that rapport is tied to verbal and non-verbal alignment (Lubold et al., 2019; Tickle-Degnen and Rosenthal, 1990), prior educational research has explored relationships between rapport with agents and various forms of alignment such as lexical (Rosenthal-von der Pütten et al., 2016; Lubold, 2018) and acoustic-prosodic (Lubold, 2018; Kory-Westlund and Breazeal, 2019) alignment.

While lexical alignment (the focus of this paper) in educational dialogue has been an active research area, prior studies are limited by 1) alignment measures (repetition of single words (Ai et al., 2010; Friedberg et al., 2012; Lubold, 2018) or manual annotations of semantics (Rosenthal-von der Pütten et al., 2016)) or 2) dialogue settings (they studied only dyadic interactions with an agent (Rosenthal-von der Pütten et al., 2016; Lubold, 2018; Sinclair et al., 2019), dyadic interactions between humans (Michel and Smith, 2017; Michel and Cappellini, 2019; Michel and O'Rourke, 2019; Sinclair and Schneider, 2021), or multi-party human interactions (Friedberg et al., 2012)). Multi-party interactions involving an agent remain to be explored with more sophisticated automated measures that can deal with the alignment of a sequence of words.

Therefore, we extend the past work on lexical alignment in educational dialogue in two ways. First, we view lexical alignment as initiation and repetition of shared lexical expressions, which are automatically extracted from dialogue excerpts and can consist of multiple words (Dubuisson Duplessis et al., 2021). Along with these metrics, we propose another viewpoint, *activeness*, which quantifies to what extent a speaker is involved in the establishment of shared expressions independent of their partner. Second, we investigate collaborative teaching where two learners co-teach a teachable NAO robot named Emma. We compare how individual

Proceedings of the SIGdial 2022 Conference, pages 615–622
Heriot-Watt University, Edinburgh, UK. 07-09, September, 2022

Figure 1: Screenshot of students and Emma in the H-H-R condition.

Mean (SD)	H-R (n=12)	H-H-R (n=26)
Utterances		
both speakers	174.8 (27.5)	59.1 (20.0)
student	81.3 (13.1)	27.7 (9.80)
Emma	93.5 (14.4)	31.5 (10.4)
Tokens		
both speakers	2919.0 (466.3)	1147.0 (388.1)
student	921.0 (228.7)	473.0 (227.2)
Emma	1998.0 (335.9)	673.0 (210.2)

Table 1: Descriptive statistics for H-R dialogues and the two Emma-student portions of H-H-R dialogues.

learners align with Emma and how alignment relates to rapport with her in this human-human-robot (H-H-R) setting versus in a one-on-one human-robot (H-R) setting. Although, outside of education, some researchers have also investigated H-H-R interactions (e.g., Kimoto et al., 2019), exploring alignment specifically in educational settings is useful because optimal alignment strategies differ from task to task (Dubuisson Duplessis et al., 2021). Through our comparisons, this paper provides the groundwork for designing different alignment strategies for teachable agents in the H-R and H-H-R settings.

2 Methodology

2.1 Data Collection

We recruited 40 undergraduate students from Pittsburgh, USA for an online study (due to COVID) over Zoom. Emma and the student(s) each had their own Zoom window (Figure 1) and conversed via speech. Students saw ratio word problems on a web application and taught them to Emma for 30 minutes. Each problem consisted of multiple steps, and students had to teach her step-by-step. Emma was designed to guide them by asking a question or making a statement relevant to their response even when they made a mistake. Her responses were pre-authored in Artificial Intelligence Markup Language and were selected based on pattern matching with students' utterances. All students were initially assigned to the H-H-R condition, but they were assigned to the H-R condition if their partner did not show up. We ended up with 12 students in the H-R condition and the remaining 28 in the H-H-R condition to form 14 pairs. In both conditions, students freely interacted with Emma by pressing and holding a "push to speak" button on the application. In the H-H-R condition, students were also expected to discuss the problems with their partners while teaching Emma, while, in the H-R condition, students had to keep talking to Emma without any discussions with others. An example H-H-R interaction can be found in Appendix A. We excluded one H-H-R pair from our analysis because one of the students did not talk to either Emma or the partner while working on the problems.

After teaching, learners individually answered survey questions about their perceived rapport with Emma on a six-point Likert scale, ranging from strongly disagree to strongly agree. The survey used four types of rapport measures created by Lubold (2018): general rapport measures (three items) based on the sense of connection from Gratch et al. (2007) and positivity, attention, and coordination rapport measures (four items each, twelve in total) from Sinha and Cassell (2015) and Tickle-Degnen and Rosenthal (1990). The latter twelve items had a higher Cronbach's α (.856) than the general rapport items (.839); thus, we used the average of the positivity, attention, and coordination items to create our rapport metric. The means and standard deviations of our rapport metric were 4.36 and .882 in the H-R condition, and 4.55 and .572 in the H-H-R condition. One-way ANOVA showed no effect of conditions on rapport ($F = .704, p = .407, df = 36$).

2.2 Computing Lexical Alignment

We manually transcribed all conversations, instead of using Emma's automated speech recognition, because she recorded only while students were holding the "push to speak" button. Then, because the measures of lexical alignment below are defined only for dyadic conversations, we manually identified the responder of each utterance in the H-H-R condition (see Appendix A) to split each conversation into two Emma-student dialogues and one student-student dialogue. Table 1 describes the Emma-student dialogue data. Although individuals in the H-H-R condition spoke less to Emma than in the H-R condition due to the fixed experiment duration and the dialogue split, this does not affect our measures because they are normalized by the number of shared expressions or tokens.

The quantification of lexical alignment in the dialogues[1] in this paper relies on a *shared expression*, which is "a surface text pattern at the utterance level that has been produced by both speakers in a dialogue" (Dubuisson Duplessis et al., 2017). A shared expression is initiated by speaker S when used by S first and adapted (thus established as a shared expression) by the dialogue partner later. We used the alignment measures derived from shared expressions because mathematical expressions often consist of more than one token, other existing measures compute only repetition, and these measures are shown to be predictive of educational outcomes. Our ratio problems contained fractions and decimals, which cannot be expressed by one word. Indeed, the average lengths of shared expressions were $1.47 \pm .076$ and $1.44 \pm .101$ for the H-R and H-H-R conditions, respectively. Word-based measures such as counting (Nenkova et al., 2008; Friedberg et al., 2012; Wang et al., 2014), Spearman's correlation coefficient (Huffaker et al., 2006), regression models (Reitter et al., 2006; Ward and Litman, 2007), and vocabulary overlap (Campano et al., 2014) fail to represent the alignment of phrases containing more than one word. Other measures address this issue by leveraging n-grams (Michel and Smith, 2017; Duran et al., 2019) or cross-recurrence quantification analysis (Fusaroli and Tylén, 2016) but consider only repetitions in the alignment process as opposed to the measures used in this work (Dubuisson Duplessis et al., 2017, 2021). Furthermore, Sinclair and Schneider (2021) have found these measures are correlated with learning and collaboration between human students in collaborative learning.

We employed the set of *speaker-dependent* alignment measures out of the ones proposed by Dubuisson Duplessis et al. (2017, 2021)[2]: Initiated Expression (IE) and Expression Repetition (ER). IE of speaker S (IE_S) measures orientation (i.e., (a)symmetry) in the alignment process and is defined as $\frac{\text{\# expr. initiated by } S}{\text{\# expr.}}$. In a dialogue between speakers $S1$ and $S2$, the alignment process is symmetric if IE_S1 $\approx$ IE_S2 $\approx .5$ because IE_S1 + IE_S2 $= 1$. ER of speaker S (ER_S) captures the strength of repetition and is defined as $\frac{\text{\# tokens from } S \text{ in new or existing expr.}}{\text{\# tokens from } S}$.

However, IE cannot measure asymmetry or establishment independent of another speaker because, by definition, if IE_S1 increases, IE_S2 decreases. This dependence prevents us from observing increased establishment by both speakers. Therefore, we calculated Expression Initiator Difference (IED) (Sinclair and Schneider, 2021), which is given by IED $= |$IE_S1 $-$ IE_S2$|$. In addition, we propose a new measure:

Expression Establishment by Speaker S (EE_S) measures the *activeness* of S in the alignment process in terms of the establishment of new shared expressions. It is given by EE_S $= \frac{\text{\# tokens from } S \text{ used to establish new expr.}}{\text{\# tokens from } S}$.

In the example dialogue in Appendix A, there are ten shared expressions in the Emma-StudentA dialogue split: "that", "can you", "can"[3], "convert", "the", "days to", "days", "to", "hours", and "hours?". Of those, Emma started to use three expressions that Student A reused later: "that", "can you", and "can". Thus, IE_Emma $= \frac{3}{10}$ and IE_student $= \frac{7}{10}$. These are used to compute IED in the Emma-StudentA dialogue: IED $= |\frac{3}{10} - \frac{7}{10}| = \frac{2}{5}$. ER_student means the number of tokens in Student A's turns that are taken from Emma's previous turns and therefore parts of shared expressions (these tokens are italicized in Appendix A) divided by the total number of tokens Student A spoke to Emma including punctuations. Student A spoke 33 tokens to Emma and devoted four italicized tokens—"can you" and two "that"s—to shared expressions. Thus, for Student A, ER_student $= \frac{4}{33}$. Out of the four, Student A used three tokens to establish new shared expressions "that" and "can you", so EE_student $= \frac{3}{33} = \frac{1}{11}$.

2.3 Alignment Hypotheses

This study investigates the following hypotheses:

H1: Individuals in the H-H-R condition align less with Emma than in the H-R condition. Bren-

[1]The original dialogues in the H-R condition and the Emma-student dialogue splits in the H-H-R condition.

[2]We used the associated tool available at `https://github.com/GuillaumeDD/dialign`.

[3]Although the expression "can" is part of the longer expression "can you", it is counted as a shared expression because it appeared as a free form in Emma's last turn (Dubuisson Duplessis et al., 2017). I.e., it was not part of "can you".

nan and Clark (1996) formulated lexical alignment as the establishment of a shared conceptualization, a conceptual pact. In the H-R condition, individuals establish conceptual pacts only with Emma, but, in the H-H-R condition, individuals do so between them through discussion before talking to Emma (see the discussion between students before talking to her in Appendix A). This may mean these conceptual pacts are likely to be different from what Emma initially suggested because humans keep updating them, but Emma is not accessible to the updated conceptual pacts (in our case, Emma does not have an ability to intentionally align with humans). Therefore, individuals in the H-H-R condition may tend to use lexicons outside of shared expressions with Emma.

H2: Students feel more rapport with Emma when they align with Emma more (H2-a), she aligns with them more (H2-b), and alignment is more symmetric (H2-c). Human-human interactions show positive correlations between alignment and rapport (Lubold et al., 2019; Tickle-Degnen and Rosenthal, 1990; Sinha and Cassell, 2015). These are bi-directional; people feel a rapport when aligning with their partners and being aligned by their partners (Chartrand and van Baaren, 2009). In human-robot interactions, positive relationships between rapport and non-lexical alignments such as acoustic-prosodic (Lubold, 2018; Kory-Westlund and Breazeal, 2019) and movement (Choi et al., 2017) have also been found. We thus expect lexical alignment positively correlates with rapport in both conditions. We also anticipate a symmetric alignment process positively correlates with rapport because human-human interactions are more symmetric than human-agent ones (Dubuisson Duplessis et al., 2021) and past work increased rapport by imitating human alignment behavior.

H3: Lexical alignment is more strongly correlated with rapport with Emma in the H-R condition than in the H-H-R condition. As shown in Yu et al. (2019), Levitan et al. (2012), and Namy et al. (2002), the alignment process in H-H-R dialogues may also depend on other factors including the gender diversity of the party. Thus, in the H-H-R condition, lexical alignment alone may not be as predictive of rapport as in the H-R condition.

3 Results and Discussion

Individual alignment across H-R and H-H-R conditions (H1). We tested H1 by comparing means of ER_student and EE_student across conditions with one-way ANOVA. Table 2 partly supports H1. Individuals in the H-R condition repeated shared expressions (i.e., higher ER_student) more than in the H-H-R condition, but they were equally likely to establish shared expressions (i.e., no difference in EE_student) across conditions.

Mean (SD)	H-R (n=12)	H-H-R (n=26)
ER_student**	.594 (.052)	.462 (.077)
EE_student	.189 (.033)	.171 (.050)
ER_Emma**	.494 (.031)	.421 (.079)
EE_Emma*	.087 (.024)	.119 (.037)
IED	.155 (.111)	.158 (.127)

Table 2: Descriptive statistics of lexical alignment measures. Measures marked with * and ** are significantly different across conditions at $p < .05$ and $p < .01$ (two-tailed), respectively.

Correlations of alignment with rapport across conditions (H2 and H3). To test H2 and H3, first, we fit the regression equation with an interaction between the conditions and an alignment measure: $R = \beta_0 + \beta_1 * HHR + \beta_2 * A + \beta_3 * HHR * A$ where R is the rapport measure, A is an alignment measure, and HHR is 1 for students in the H-H-R condition; otherwise 0. Table 3 shows that β_3 is not significant for none of the alignment measures, meaning that the correlations between rapport and alignment are in the same direction regardless of the conditions.

Therefore, we used all data to compute Pearson's correlations between rapport and alignment (see Table 4). The significant negative correlation between rapport and IED supports H2-c. H2-b is not fully supported because, although EE_Emma is correlated positively with rapport, ER_Emma is not. In addition, surprisingly, we found evidence for the opposite of H2-a; EE_student has a negative correlation with rapport. Further analysis revealed IE_Emma is significantly negatively correlated with rapport ($r = -.490, p = .002$). This means students felt less rapport when they established more shared expressions relative to Emma and aligns with the findings on EE.

Finally, we compared Pearson's r between lexical alignment and rapport in the two conditions using Fisher transformation (Snedecor and Cochran, 1980) to test H3. It was not validated because there was no significant difference between the two conditions in Table 5.

Estimate of β_3 (p-value)	ER_student	EE_student	ER_Emma	EE_Emma	IED
Rapport	-0.54 (.901)	9.09 (.171)	3.10 (.652)	-0.83 (.928)	3.72 (.057)

Table 3: Coefficients of interaction terms (β_3).

Pearson's r (p-value)	ER_student	EE_student	ER_Emma	EE_Emma	IED
Rapport	-.315 (.054)	-.331* (.043)	.214 (.198)	.343* (.035)	-.573** (.000)

Table 4: Pearson's correlations between alignment measures and rapport. Correlations marked with * and ** are significant at $p < .05$ and $p < .01$ (2-tailed), respectively.

Pearson's r	H-R (n=12)	H-H-R (n=26)
ER_student	-.145	-.406
EE_student	-.457	-.285
ER_Emma	.008	.461
EE_Emma	.195	.407
IED	-.723	-.529

Table 5: Comparison of Pearson's correlations between alignment measures and rapport across conditions.

These results may be because perceived success in communication with Emma characterized by her accidental alignment leads to high rapport and low alignment by students. As Branigan et al. (2010) and Dubuisson Duplessis et al. (2017) reported, students might have (either consciously or unconsciously) expected they should establish shared expressions more than Emma due to her limited linguistic capacity. Thus, they might have started with an asymmetric alignment process. When Emma was stuck, they might have kept this strategy because they thought she did not understand them, resulting in decreased rapport. In contrast, as Emma established new shared expressions by accident, students might have thought she was following new information like humans, that she cared what they said, and that they were in sync, leading to more positivity, attention, and coordination rapport, respectively (Tickle-Degnen and Rosenthal, 1990). They may have also changed their alignment strategy to a more symmetric one (i.e., decreased alignment by students) that they usually use while interacting with humans.

3.1 Limitations

This study has several limitations. First, the limited number of participants (38 in total) might limit the detection of all correlations. Moreover, the comparison between the H-R and H-H-R conditions has low statistical power because the H-R condition had fewer than half of the participants in the H-H-R condition. It might have been biased because the assignment to the conditions was not fully random as well. Next, alignment measures may need contextual adjustments. For example, one math problem included both "three hours" and "three-fortieths of battery". Although "three" in these numbers refers to different entities, our measures saw it as a shared expression. Finally, some lexicons came from the problem prompt rather than the group conversation.

4 Conclusion

We examined relationships between lexical alignment and rapport with a teachable agent in one-on-one (H-R) and collaborative (H-H-R) teaching. Our methods expand prior literature by comparing alignment behavior in H-R and H-H-R settings and extending recent work by Dubuisson Duplessis et al. (2021) to the speaker-level act of *activeness* in the alignment process. Our results imply learners' lexical alignment with teachable agents may not always increase rapport with a teachable agent, unlike predictions from alignment theories (Lubold et al., 2019; Tickle-Degnen and Rosenthal, 1990) largely based on human-human interactions. Future work can expand our work by looking at the role of H-H portions of H-H-R interactions in their H-R portion and the effect of miscommunication as an intermediate variable on the negative correlations between rapport and learners' alignment and by extending the measures to multi-party settings without disentanglement.

Acknowledgements

We would like to thank anonymous reviewers for their thoughtful comments on this paper. This work was supported by Grant No. 2024645 from the National Science Foundation, Grant No. 220020483 from the James S. McDonnell Foundation, and a University of Pittsburgh Learning Research and Development Center internal award.

References

Hua Ai, Rohit Kumar, Dong Nguyen, Amrut Nagasunder, and Carolyn P. Rosé. 2010. Exploring the effectiveness of social capabilities and goal alignment in computer supported collaborative learning. In *Proceedings of the 10th International Conference on Intelligent Tutoring Systems - Volume Part II*, ITS'10, page 134–143, Berlin, Heidelberg. Springer-Verlag.

Holly P. Branigan, Martin J. Pickering, Jamie Pearson, and Janet F. McLean. 2010. Linguistic alignment between people and computers. *Journal of Pragmatics*, 42(9):2355–2368. How people talk to Robots and Computers.

Susan E. Brennan and Herbert H. Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22:1482–1493.

Zoraida Callejas, David Griol, and Ramón López-Cózar. 2011. Predicting user mental states in spoken dialogue systems. *EURASIP Journal on Advances in Signal Processing*, 2011(1):1–21.

Sabrina Campano, Jessica Durand, and Chloé Clavel. 2014. Comparative analysis of verbal alignment in human-human and human-agent interactions. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 4415–4422, Reykjavik, Iceland. European Language Resources Association (ELRA).

Sabrina Campano, Caroline Langlet, Nadine Glas, Chloé Clavel, and Catherine Pelachaud. 2015. An eca expressing appreciations. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 962–967.

Tanya L. Chartrand and Rick van Baaren. 2009. Chapter 5 human mimicry. volume 41 of *Advances in Experimental Social Psychology*, pages 219–274. Academic Press.

Mina Choi, Rachel Kornfield, Leila Takayama, and Bilge Mutlu. 2017. Movement matters: Effects of motion and mimicry on perception of similarity and closeness in robot-mediated communication. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 325–335, New York, NY, USA. Association for Computing Machinery.

Guillaume Dubuisson Duplessis, Chloé Clavel, and Frédéric Landragin. 2017. Automatic measures to characterise verbal alignment in human-agent interaction. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pages 71–81, Saarbrücken, Germany. Association for Computational Linguistics.

Guillaume Dubuisson Duplessis, Caroline Langlet, Chloé Clavel, and Frédéric Landragin. 2021. Towards alignment strategies in human-agent interactions based on measures of lexical repetitions. *Language Resources and Evaluation*, 55(2):353–388.

Nicholas Duran, Alexandra Paxton, and Riccardo Fusaroli. 2019. Align: Analyzing linguistic interactions with generalizable tech niques-a python library. *Psychological Methods*, 24(4):419–438.

Heather Friedberg, Diane Litman, and Susannah B. F. Paletz. 2012. Lexical entrainment and success in student engineering groups. In *2012 IEEE Spoken Language Technology Workshop (SLT)*, pages 404–409.

Riccardo Fusaroli and Kristian Tylén. 2016. Investigating conversational dynamics: Interactive alignment, interpersonal synergy, and collective task performance. *Cognitive science*, 40(1):145–171.

Jonathan Gratch, Ning Wang, Jillian Gerten, Edward Fast, and Robin Duffy. 2007. Creating rapport with virtual agents. In *Intelligent Virtual Agents*, pages 125–138, Berlin, Heidelberg. Springer Berlin Heidelberg.

Agneta Gulz, Magnus Haake, and Annika Silvervarg. 2011. Extending a teachable agent with a social conversation module: Effects on student experiences and learning. In *Proceedings of the 15th International Conference on Artificial Intelligence in Education*, AIED'11, page 106–114, Berlin, Heidelberg. Springer-Verlag.

David Huffaker, Joseph Jorgensen, Francisco Iacobelli, Paul Tepper, and Justine Cassell. 2006. Computational measures for language similarity across time in online communities. In *Proceedings of the Analyzing Conversations in Text and Speech*, pages 15–22, New York City, New York. Association for Computational Linguistics.

Mitsuhiko Kimoto, Takamasa Iio, Michita Imai, and Masahiro Shiomi. 2019. Lexical entrainment in multi-party human–robot interaction. In *Social Robotics*, pages 165–175, Cham. Springer International Publishing.

Jacqueline M. Kory-Westlund and Cynthia Breazeal. 2019. Exploring the effects of a social robot's speech entrainment and backstory on young children's emotion, rapport, relationship, and learning. *Frontiers in Robotics and AI*, 6.

Krittaya Leelawong and Gautam Biswas. 2008. Designing learning by teaching agents: The betty's brain system. *Int. J. Artif. Intell. Ed.*, 18(3):181–208.

Rivka Levitan, Agustín Gravano, Laura Willson, Štefan Beňuš, Julia Hirschberg, and Ani Nenkova. 2012. Acoustic-prosodic entrainment and social behavior. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 11–19, Montréal, Canada. Association for Computational Linguistics.

Nichola Lubold. 2018. *Producing Acoustic-Prosodic Entrainment in a Robotic Learning Companion to Build Learner Rapport*. Ph.D. thesis, Arizona State University.

Nichola Lubold, Erin Walker, Heather Pon-Barry, and Amy Ogan. 2018. Automated pitch convergence improves learning in a social, teachable robot for middle school mathematics. In *Artificial Intelligence in Education*, pages 282–296, Cham. Springer International Publishing.

Nichola Lubold, Erin Walker, Heather Pon-Barry, and Amy Ogan. 2019. Comfort with robots influences rapport with a social, entraining teachable robot. In *Artificial Intelligence in Education*, pages 231–243, Cham. Springer International Publishing.

Marije Michel and Marco Cappellini. 2019. Alignment during synchronous video versus written chat l2 interactions: A methodological exploration. *Annual Review of Applied Linguistics*, 39:189–216.

Marije Michel and Breffni O'Rourke. 2019. What drives alignment during text chat with a peer vs. a tutor? insights from cued interviews and eye-tracking. *System*, 83:50–63. Special Edition on the Work of Professor Stephen Bax.

Marije Michel and Bryan Smith. 2017. *Measuring lexical alignment during L2 chat interaction: An eye-tracking study*, pages 244–267. Taylor and Francis.

Laura L. Namy, Lynne C. Nygaard, and Denise Sauerteig. 2002. Gender differences in vocal accommodation:: The role of perception. *Journal of Language and Social Psychology*, 21(4):422–432.

Ani Nenkova, Agustín Gravano, and Julia Hirschberg. 2008. High frequency word entrainment in spoken dialogue. In *Proceedings of ACL-08: HLT, Short Papers*, pages 169–172, Columbus, Ohio. Association for Computational Linguistics.

David Reitter, Frank Keller, and Johanna D. Moore. 2006. Computational modelling of structural priming in dialogue. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, pages 121–124, New York City, USA. Association for Computational Linguistics.

Astrid M. Rosenthal-von der Pütten, Carolin Straßmann, and Nicole C. Krämer. 2016. Robots or agents – neither helps you more or less during second language acquisition. In *Intelligent Virtual Agents*, pages 256–268, Cham. Springer International Publishing.

Arabella Sinclair, Kate McCurdy, Christopher G Lucas, Adam Lopez, and Dragan Gašević. 2019. Tutorbot corpus: Evidence of human-agent verbal alignment in second language learner dialogues. *International Educational Data Mining Society*.

Arabella J Sinclair and Bertrand Schneider. 2021. Linguistic and gestural coordination: Do learners converge in collaborative dialogue?. *International Educational Data Mining Society*.

Tanmay Sinha and Justine Cassell. 2015. We click, we align, we learn: Impact of influence and convergence processes on student learning and rapport building. In *Proceedings of the 1st Workshop on Modeling INTERPERsonal SynchrONy And InfLuence*, INTERPERSONAL '15, page 13–20, New York, NY, USA. Association for Computing Machinery.

G.W. Snedecor and W.G. Cochran. 1980. *Statistical Methods*, seventh edition. Iowa State University Press.

Linda Tickle-Degnen and Robert Rosenthal. 1990. The nature of rapport and its nonverbal correlates. *Psychological Inquiry*, 1(4):285–293.

Yafei Wang, David Reitter, and John Yen. 2014. A model to qualify the linguistic adaptation phenomenon in online conversation threads: Analyzing priming effect in online health community. In *Proceedings of the Fifth Workshop on Cognitive Modeling and Computational Linguistics*, pages 55–62, Baltimore, Maryland, USA. Association for Computational Linguistics.

Arthur Ward and Diane J Litman. 2007. Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora.

Mingzhi Yu, Diane Litman, and Susannah Paletz. 2019. Investigating the relationship between multi-party linguistic entrainment, team characteristics and the perception of team social outcomes. In *The Thirty-Second International Flairs Conference*.

A Sample Dialogue and Lexical Alignment

Speaker	Utterance	Responder
Emma:	**Now that I know how long one battery will last, can you help me figure out how many batteries I need total?**	**Student A**
Student B:	Oh, okay.	Student A
Student A:	Okay. Next ...	Student B
Student B:	Okay.	Student A
Student A:	So one whole battery lasts three and three quarters of an hour.	Student B
Student B:	Three and three quarters of an hour. Oh my gosh. Emma, you're making this difficult on us.	Student A
Student A:	I think we need her... Oh man.	Student B
Student B	Because it's in days now. So she has to figure out...	Student A
Student A:	You have to do like dimensional analysis.	Student B
Student B	I think she... Yeah. She has to convert days to hours. I think that might be the easiest thing for her.	Student A
Student A:	Okay.	Student B
Student B:	So she has to divide...	Student A
Student A:	Wait, is she going to remember that?	Student B
Student B:	Oh, I don't know.	Student A
Student A:	Okay. I'm going to ask if she knows how to convert days to hours.	Student B
Student B:	Okay.	Student A
Student A:	**Okay, Emma. *Can you* convert the number of days to the number of hours?**	**Emma**
Emma:	**So I know how long I'll be gone in *days*, but how long *the* battery lasts is in *hours*. So first I should change *the days to hours?***	**Student A**
Student A:	**Yes, Emma. *That*'s correct.**	**Emma**
Emma:	**So I *can convert* two and three quarters *to* an improper fraction, eleven over four. And then I *can* multiply it by twenty four *hours?***	**Student A**
Student B:	She did all the work.	Student A
Student A:	**Yes, Emma. *That*'s correct.**	**Emma**

Table 6: Lexical alignment in the Emma-studentA portion of the dialogue (bolded utterances). A responder is a speaker who responded to the utterance. Speakers initiated colored but not italicized expressions and repeated the italicized ones. Contractions were tokenized as two tokens (e.g., That's to "That" and "'s"). Punctuation was treated as one token but did not constitute a shared expression by itself.