Online Meetings: When Repetition Signals Engagement

Anaïs Claire Murat
School of Computer Science and Statistics
Trinity College Dublin
Dublin, Ireland
murata@tcd.ie

Carl Vogel
School of Computer Science and Statistics
Trinity College Dublin
Dublin, Ireland
vogel@cs.tcd.ie

***Index Terms*—interaction, online meeting, online tutoring, engagement, lexical spread, entropy, repetition**

***Abstract*—Online meetings are challenging environments to assess and foster engagement. By observing the relationship between engagement and lexical spread in the RoomReader corpus, a dataset that records the engagement of 3-4 participants during online sessions moderated by a tutor, we find evidence that dialogue participants' engagement increases with the tutor's back-channeling whereas the participants' production associates with significantly greater engagement when it reuses words from the previous turns and builds on ideas.**

I. Introduction

Observation of communicative interactions frequently includes assessment of engagement of the participants. Participants have an interest in whether their interlocutors are engaged in their interaction or distracted; external observers may decide whether to join in or to interrupt an interaction on the basis of how engaged participants appear to be. It is interesting to explore the behaviours that associate with perceptions of engagement, and natural to look to properties of the content of communication as well as non-linguistic behaviours such as gaze and other bodily motions. Where communication is conducted online, the relation between gaze and engagement perceptions is more difficult to assess, but one may, as we do here, study abstract features of communication content. We focus on lexical unigram other-repetition and forms of entropy as linguistic indicators, and study the relationship between these and third-party assessments of engagement in a multimodal corpus of online synchronous communications. Our hypothesis is that engagement is dependent on the conversation unfolding in ways that can be perceived through the study of lexical spread. We make this more precise in what follows.

Cognitive infocommunications is concerned with the enhancement of human cognitive capabilities [1]–[4]. Understanding the status quo in human cognitive capabilities remains fundamental to this pursuit [5]. Linguistic and behavioural interaction analysis is a constant thread of research within cognitive infocommunications [6]–[11]. Virtual reality environments are included in the interaction contexts addressed [12], [13], and synchronous online communication is a species of virtual reality. Engagement is as relevant to virtual reality as face to face interactions, and reliable indicators of engagement possibly even more necessary in virtual settings.

After looking into the literature about engagement (§II) and describing our method to qualify lexical spread (§III), we discuss results (§IV) suggesting that (1) engagement is greater for turns reusing vocabulary – and to a more refined level for the tutor's production than for the participants'; and that, more specifically, (2) minimal validation, backchannelling and providing feedback are correlated with greater engagement. We conclude (§V) with, among other remarks, future directions.

II. Background Literature

Engagement is valuable in many domains, from teaching to sales, through enhancing productivity in the workplace. As such, the notion of engagement has suffered from vagueness. However, current literature in education studies [14], [15] recognises and values three interconnected dimensions of engagement: cognitive (whether one is actually learning), emotional (related to the feeling experienced towards the institution), and behavioral (what actions are undertaken, including active participation, attention and outside-the-interaction actions) [16]. The challenges accompanying engagement are at least two-fold: one needs to (1) be able to assess engagement to then (2) foster it. Many have looked into assessing engagement through task success [17], [18], gestures [19], [20], emotion recognition [19], and eye-tracking [19], [21], [22]. In the context of online education, fostering engagement has been studied through a spectrum of collaboration tools [23], [24] or communication tools [25], including generative AI [26].

Online meetings enable instantaneous dialogue, but they differ from face-to-face interactions in that they come with fatigue [27] notably due to overcompensation for non-verbal cues and distorted eye-gaze [28]. Little work has focused exactly on the content of how things are said in relation to engagement, even though dialogue in itself can be a *tool* to reach engagement. [15] highlights the required properties of a dialogue for it to lead to an engaged audience. Specifically, in organization–public interactions, they emphasize the role of the initiator as a participant who should promote the other's participation and value their contribution. Doing so requires listening skills and being able to build upon one's production.

This publication has emanated from research conducted with the financial support of Taighde Éireann-Research Ireland under Grant. No. 18/CRT/6223.

[29] also highlights the importance of building a tutor-student bond, notably through sharing experiences.

This work examines the spoken content of online sessions, providing insights into assessing engagement and suggesting linguistic tools to improve engagement during online meetings, especially when a facilitator-participant hierarchy is present.

III. Method

We deploy observational methods in this work. In the spirit of such methods as defended elsewhere [6], our goal is to notice regularities and irregularities, for their own sake and for the sake of informing broader theories. Here we provide more details about the corpus of conversations that we analyze (its pre-processing) and the measurements that we record over the course of the conversations and analyze in aggregate.

A. Dataset & Engagement: The RoomReader Corpus

The RoomReader corpus was designed to enable the measurement of engagement in tutor-students-like online sessions. It accounts for 30 recorded conversations, including between 2 and 4 participants (118 participants total) and a tutor (2 tutors in total). While the tutor behaves as a facilitator, participants need to collaborate to find the most popular answers to prompts like "name something that cuts", similar to the TV show Family Freud. On top of collecting post-session self-reports of engagement, independent annotators also annotated the continuous engagement of the students based on the Zoom recordings [30], hence providing insights into the conversation dynamics [31]. Continuous engagement was tracked every second on a scale from -2 to 2 where -2 represents a participant lying back and not paying attention, and 2 is a participant speaking and contributing to the session [18]. In this study, to analyze the relationship between the two modalities that are turns and engagement, we used the time-stamps provided with each of the annotations to first average each participant's engagement (E_p) measure over the length of the turn, and then compute a "total engagement" (E_{total}) such as eq. 1.

$$E_{total} = \frac{E_{p_1} + E_{p_2} + \cdots + E_{p_n}}{n} \tag{1}$$

B. Pre-processing of the transcripts

For the first part of our study processing repetitions and entropy, the RoomReader transcripts were normalised: punctuation was removed, all letters were made lowercase, all "pronoun + verb" contractions were split (e.g "you're" into "you", "are"), and all auxiliaries and negations were equally split (e.g "can't" into "can", "not"). Note that only the lowercase switch was kept for the post-hoc token analysis since punctuation was useful for providing insights into the quality of turns (see "yeah." vs "yeah," or "that's" vs "that" vs "is" in Tab. II).

C. Lexical Repetition

As [15], [20], [29] highlighted, engaging conversations require respect and this comes through attentive listening and value for each interlocutor's contribution. In this spirit, we draw a particular focus on keeping track of other-repetitions (OR): when one is able to reuse someone else's words. Lexical repetitions in natural language partly depend on a text's tokenization, but the literature on alignment does not adopt a single view on such tokenization. Options include fully inflected lexical forms, lemmatized forms, multiple-word-expression individuation, pairings of lemmas with a part-of-speech, classification of its use in context, and so on. Differences also emerge from looking into the type of repetition (lexical vs syntactic repetition [32]), and into the calculation models (normalised frequencies, strict counts, etc.) [33]. To address some of these distinctions and to avoid mere speculation on the type of content that is being repeated, we made a distiction between Open-Class (OC) repetitions, and Closed-Class (CC) repetitions; the latter allowing more syntactic repetition. To do so, we supplemented the existing corpus with Part-of-Speech (POS) tags using SpaCy pre-trained small model for English. This implies that to account for a repetition, POS and tokens have to be the same. Based on these tags, each word is then categorised as "OC" if belonging to the proper noun, noun, verb, or adjective classes, or "CC" if else. Assessing the overall similarities of all interlocutor's productions is not enough: it does not allow to draw any conclusion on the impact of what one previously said on the current production [34]. Temporal cues are therefore crucial and we do so by comparing strictly ordered turns [35], [36].

Linguistic repetitions are thus calculated at the turn level as a Jaccard Index in which are compared the number of tokens (w) of a turn (T_n) being repeated over the types (v) from the previous turn (T_{n-1}) that could have been repeated but that were not, such that defining (2) and (3),[1]

(2) $T_n = \{w_1, w_2, ..., w_i\}^M$ (where i is the number of *tokens* (w) in a given turn, and ^{M}emphasizes its multiset characterisation),

(3) $T_{n-1} = \{v_1, v_2, ..., v_j\}$ (where j is the number of *types* (v) in the previous turn),

gives Jaccard index calculation $J(T_n, T_{n-1})$ in (7):[2]

$$J(T_n, T_{n-1}) = \frac{|T_n \cap T_{n-1}|}{|T_n \cup T_{n-1}|} \tag{7}$$

For example, if one says "Hello hello, I'm good" in response to "Hello you, how are you?", we would obtain the following: $T_n = \{hello, hello, i, am, good\}^M$ and $T_{n-1} = \{hello, you, how, are\}$. Note $|T_n \cap T_{n-1}| = \{$hello, hello$\}$. So the Jaccard index J attributed to T_n is: $2 \div 7$.

This turn by turn manner of counting repetition does not account for repetition on the overall conversation level. Hence,

[1] Notice the intentional asymmetry in counting *types* and *tokens*.

[2] We assume a commutative intersection operator between a set Σ and multiset Σ^M which we define as the smallest multiset satisfying (4); the union of a set Σ and multiset Σ^M is the smallest multiset satisfying (5) and (6).

$$\text{for each } x \in \Sigma^M, \text{ if } x \in \Sigma \text{ then } x \in \Sigma \cap \Sigma^M \tag{4}$$

$$\text{for each } x \in \Sigma^M \text{ if } x \notin \Sigma \text{ then } x \in \Sigma \cup \Sigma^M \tag{5}$$

$$\text{for each } x \in \Sigma \text{ then } x \in \Sigma \cup \Sigma^M \tag{6}$$

we complement this method with entropy measures that keep track of all the words already uttered across the dialogue.

D. Entropy

Entropy quantifies the information required to encode a message [37]. The more complex a message is, the more information is required to encode it; the higher the entropy is. [38] aimed to detect critical changes in a dialogue's flow by measuring entropy over strings of dialogue acts. We adapt their method directly to tokens to assess lexical development.

Let $V = \{v_1, v_2, ..., v_k\}$ be all the item types v available in a given dialogue D. D can be chronologically defined as a sequence of turns T such as $D = T_1, T_2, ..., T_n$ (abbreviated T_1^n) which are themselves defined as sets T of types such as $T = w_l, ..., w_m$ where l and m are respectively the index of the first and last type[3] of the turn T_n, and $(\forall 1 \leq j \leq m)(w_j \in V)$. The entropy H of any dialogue from its start to its nth turn (T_1^n) is based on Shannon Entropy such that:

$$H(T_1^m) = -\sum_{1}^{k} P(v_i)\log_e P(v_i) \tag{8}$$

where probability $P(v_i)$ is the relative frequency of the type w_i throughout the observed dialogue sequence.

To observe the impact of one turn over the entropy of the dialogue, one can compare the entropy at the end of a given turn h_n to the entropy at the end of the previous turn h_{n-1}:

$$\Delta H = \Delta(h_n - h_{n-1}) \tag{9}$$

IV. Results & Discussion

This observational study began with the hypothesis that engagement is dependent on the conversation unfolding in ways that can be perceived through the study of lexical spread. As detailed earlier, entropy in this study has been designed as a means of exploring the unfolding of the conversation, providing a quantitative method to measure vocabulary use and reuse within a given conversation.

TABLE I
Statistical comparison of Engagement between $\Delta H > 0$ and $\Delta H < 0$ (Wilcoxon rank sum test, p-value $< 2.2e - 16$)

Statistic	$\Delta H > 0$	$\Delta H < 0$
Min	-0.24	-0.24
Max	1.88	2.00
Mean	1.19	1.23
Median	1.19	1.23
Standard Deviation	0.24	0.20

In the overall dataset, engagement is greater when $\Delta H < 0$ (Tab. I) (Wilcoxon test, $p < 2.2e^{-16}$) [4]. This suggests that the participants' engagement is greater in moments of redundancy of vocabulary. Such moments can indicate either grounding through reuse of contentful lexical terms or backchanelling ("uhm", "yeah", etc.). Thus, we examine subcategories of turns based on their repetition qualities for each locutor: do they contain other-repetitions (OR)? If so, do these include repetitions of open-class (OC) or closed-class (CC) words?

Fig. 1 (a) shows, by pair — with and without other-repetition — the difference of engagement based on the increase or decrease in entropy. In each pair, the tutors' turns which led to a decrease in entropy have, on average, significantly greater engagement ($p = 3.04e^{-14}$ for pairs without repetitions, and $p = 1.96e^{-04}$ for pairs with repetition). Furthermore, the subcategory of turns which lead to an increase in entropy and do not display any OR have significantly lower engagement from all the other combinations ($p = 1.96e^{-18}$ against $\Delta H > 0$, OR; $p = 7.27e^{-31}$ against $\Delta H < 0$, OR). On the other hand, looking at participants' productions (Fig. 1 (b)), while $\Delta H > 0$ with no OR is still leading to significantly less engagement overall ($p = 2.79e^{-03}$ against $\Delta H < 0$ without OR, $p = 6.02e^{-13}$ against $\Delta H < 0$ with OR, and $p = 5.67e{-}09$ against ΔH with OR), the difference in engagement is not significant in the pair containing OR. Hence, while the distinction between $\Delta H < 0$ and $\Delta H > 0$ remains present for the non-repeating turns, the difference between the subcategories might be mostly explained by OR presence or absence, OR subcategories showing more engagement, independently of the impact on entropy ($p = 0.02$ in $\Delta H < 0$ pairs, and $p = 5.67e{-}09$ in $\Delta H > 0$ pairs).

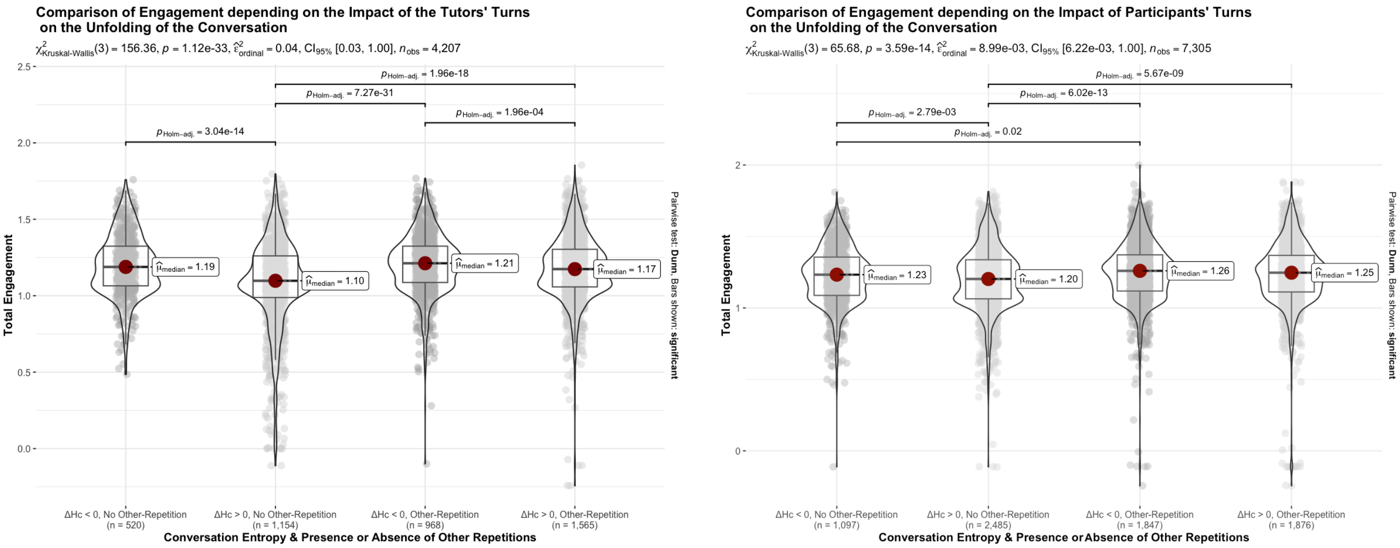
These results seem to indicate a difference of impact in the participants and the tutors' production in the RoomReader corpus. $\Delta H < 0$ triggering more engagement requires paying attention in length to the conversation, whereas OR only require a memory span of two turns. Hence, these results showing greater impact of ΔH for the engagement in the case of turns by the tutor than the participants might indicate that participants keep track of the tutor's statements better than theirs as a whole, and, in order for the engagement of the participants to significantly increase during participants' turns, it is necessary for them to make their statement obviously related to previous turn by using direct OC and CC repetitions.

More refined subcategories accounting for the type of repetition (OC or CC) are available in Fig. 2 to better understand what OR implies. Fig. 2 (a) shows the tutor's turns in relation to the total engagement. Unlike Fig. 1 (a), pairs of repetitions are not all significantly different based on ΔH . In fact, only pairs which do not account for OC repetitions are ($p = 1.75e^{-13}$ for pairs with no repetitions, and $p = 7.90e^{-14}$ for pairs with CC repetitions only). As for participant's productions (Fig. 2 (b)), ΔH is only discriminative inside the pair of turns including no repetitions ($p = 0.01$). Furthermore, turns with no repetition leading to a $\Delta H > 0$ are also different from $\Delta H > 0$ turns including OC repetitions. Overall, it seems that turns which account for OC and CC repetitions do get greater engagement than turns with no repetition at all ($p = 1.15e^{-03}$ for $\Delta H < 0$ and $p = 8.12e^{-09}$ for $\Delta H > 0$), or only CC repetitions and $\Delta H > 0$ ($p = 0.01$ for $\Delta H < 0$, OC-CC, and $p = 4.79e^{-04}$ for $\Delta H > 0$, OC-CC).

In the case of the tutor, ΔH can be significantly correlated

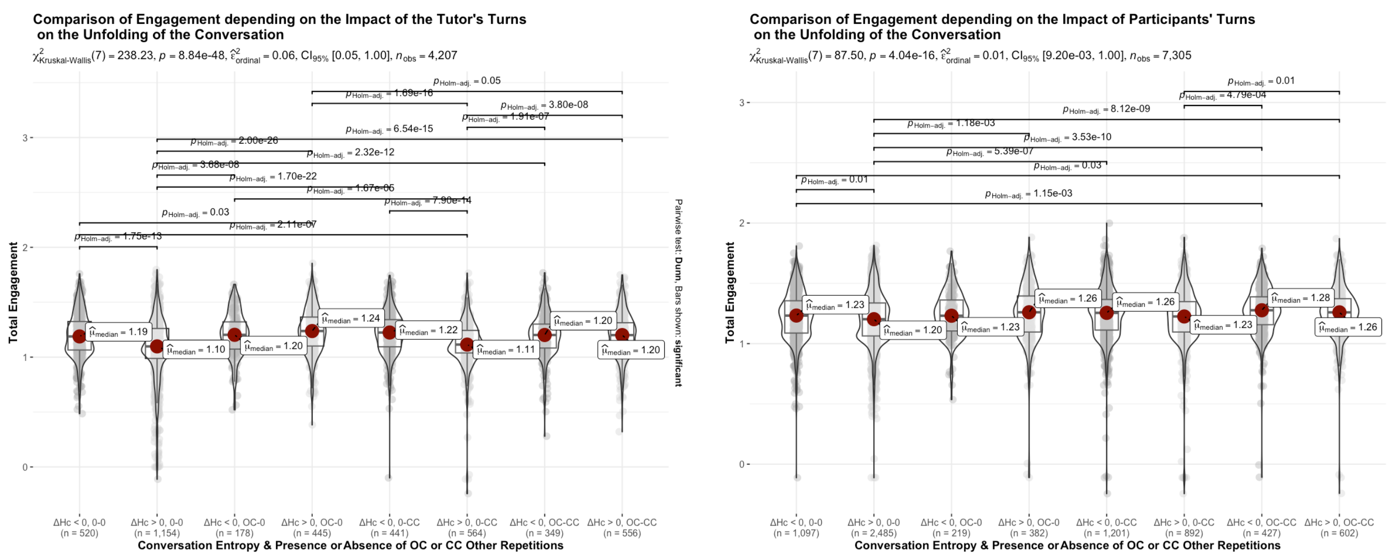
[3] First and last by an arbitrary ordering of the set, say alphabetic.

[4] Engagement was not normally distributed (Kolmogorov-Smirnov test, $p < 2.2e^{-16}$), hence, all statistical tests that follow are non-parametric.



(a) (b)

Fig. 1. Engagement Quantification depending on the Conversation Entropy Difference from one turn to another and the Presence or Absence of OR



(a) (b)

Fig. 2. Engagement Quantification depending on the Conversation Entropy Difference from one turn to another and the Presence or Absence of Open-Class and/or Closed-Class OR

with engagement without OC repetitions. This implies that turns correlated with higher engagement do not necessarily directly rephrase or reuse items from adjacent turns. The following exploration is thus rather qualitative, to better explain what types of tutor turns lead to greater or lesser engagement. We do so by assessing the importance of specific tokens to a certain entropy category: we draw one contingency table for every token x available in the corpus ($n = 5216$). A χ^2 test is then applied, and we disregard any word that led to any of the expected values to be less than 5. As we intend to inspect all of the adjusted standardized residuals to identify items of interest, we adjust $\alpha = 0.05$ by the number of tests ($n = 5216$), $\alpha' = 9.59e^{-6}$, and use the cutoff of $N(0,1)^{\alpha'} = 4.27$ for determining significance. Tab. II shows the tokens for which the χ^2 standard residuals (r) were significant in appearance likelihood based on the ΔH and repetition context. Words associated with $\Delta H > 0$ and no repetitions include “you”, “going”, “to”, “name”, “something”, “most”, “people”, etc, ($r \geq 5.7$) unlike any other categories, making it clear that this context tends to host tutors’ instructions. Similarly, $\Delta H > 0$ with CC repetitions are related with “or” ($r = 5.5$) which in this category is mostly associated with the second phase of the quiz (“which of x, y, *or* z ranks first?”). On the opposite, in the absence of repetition, $\Delta H < 0$ seems to

TABLE II

STANDARD RESIDUALS (r) OF THE CHI-SQUARE TESTS HIGHLIGHTING THE RELEVANCE OF A WORD TO ENTROPY AND REPETITION CONTEXTS OF THE **TUTOR**'S TURNS. *Bold = significant residual*

	ΔH <0				ΔH >0			
Token	**ø**	**CC**	**OC**	**OC-CC**	**OC-CC**	**OC**	**CC**	**ø**
name	-3.0	-2.8	-2.3	-2.8	-2.9	-1.4	1.4	**8.0**
most	0.3	0.9	-0.7	-0.3	-2.6	-2.5	**-4.3**	**7.6**
to	-0.9	0.8	-2.0	-2.4	-3.2	-3.8	0.2	**7.6**
people	-2.7	-2.4	-3.0	-1.8	-2.1	0.3	-0.4	**7.1**
you	1.4	-1.5	-2.4	-2.7	-3.4	-1.8	0.3	**6.8**
going	-0.6	-0.6	-0.9	0.5	-2.6	-3.0	-1.5	**6.2**
some-thing	-1.2	-0.4	-1.6	-1.7	-2.1	-1.4	-0.4	**5.7**
if	-2.0	-1.9	-2.4	-1.6	-2.0	-1.5	2.2	**4.9**
or	-2.9	-2.2	-2.1	-2.0	0.2	-1.7	**5.5**	0.8
that	-1.5	1.6	-2.6	1.6	-1.6	-3.8	**4.5**	-0.3
$	**7.4**	2.7	2.5	1.5	-1.0	-1.8	-1.7	-3.4
there.	2.0	**4.4**	3.2	0.8	-1.2	4.2	-3.2	-3.9
no,	3.0	2.2	-0.6	3.7	-1.1	**4.3**	-3.0	-3.9
on	-1.2	2.3	2.9	1.0	1.2	2.9	-1.3	**-4.3**
like	-3.2	3.6	-2.7	0.9	**4.5**	**-4.3**	3.9	**-4.8**
is	-1.0	-1.7	**6.6**	3.2	1.1	**4.8**	-2.7	**-4.9**
not	0.4	3.2	**4.4**	3.8	-1.2	**8.9**	**-4.9**	**-6.4**
one	3.3	1.4	**6.8**	**5.2**	-3.4	**4.8**	-2.1	**-6.4**
i	-0.8	3.0	-3.4	**5.6**	**5.6**	-3.0	-0.1	**-6.8**
it's	3.0	**10.4**	2.2	**7.5**	-0.2	-2.3	**-4.8**	**-6.9**
that's	**6.7**	**8.4**	**4.4**	**8.2**	-1.8	-0.1	**-4.3**	**-9.3**
yeah.	**15.8**	**30.6**	0.4	-1.7	**-6.4**	-3.1	**-6.8**	**-9.4**
yeah,	-1.3	**9.5**	2.1	**6.3**	1.3	-2.9	1.8	**-10.4**

mostly relate with minimal validation (e.g. "yeah.", $r = 15.8$, but not "yeah," $r = -1.3$, "that's" ($r = 6.7$) but not "that" ($r = -1.5$), and "$" which denotes disfluencies, particularly present after "uhm" or "uh" in this context). In the presence of CC repetitions, significant words include "yeah." ($r = 30.6$) and also "yeah," ($r = 9.5$) along with "that's" ($r = 8.4$), "it's" ($r = 10.4$) and "there." ($r = 4.4$). Sentence examples with these tokens drawn from this category include positive and negative feedback (e.g. "Yeah, that's all there fair play."). $\Delta H > 0$ with only OC repetitions is associated with "no," (e = 4.3), "is" ($r = 4.8$), "not" ($r = 8.9$),"one" ($r = 4.8$). $\Delta H < 0$ OC repetitions displays similar correlations, except for "no," and it is further correlated with "that's" ($r = 4.4$) and to a lesser extent with "it's" ($\Delta H < 0 : r = 2.2$, $\Delta H > 0 : r = -2.3$) and "yeah," ($r = 2.1$ for $\Delta H < 0$ but $r = -2.9$ for $\Delta H > 0$), making $\Delta H > 0$ with OC repetitions only a category for, almost exclusively, negative feedback. $\Delta H < 0$ and OC-CC repetitions is strongly correlated with "yeah," ($r = 6.3$), "it's", "that's", "I", "one" ($r > 5.2$). Whereas $\Delta H > 0$ OC and CC repetition is mostly defined through "I" ($r = 5.6$) and "like" ($r = 4.5$) which denotes that this category is particularly marked by the subjectivity of the tutor such as instances when they give personal opinions about the participants' answers (e.g: "I feel like that should totally be on there because I've just been doing my taxes, but it's not.")

In short, we can observe that $\Delta H > 0$ with no repetition is home to the initial instructions provided by the tutor. The second part of the task — to order the items by popularity — is particularly visible in $\Delta H > 0$ including CC repetitions. $\Delta H < 0$, in comparison, is marked by the presence of backchannelling instances (no repetitions), or moments (CC repetitions), and very brief feedback (such as "that's correct"). The presence of OC repetition usually conveys the explicit restating of the game's current situations (correct or incorrect answers already found). $\Delta H > 0$ with OC repetitions only seem to be characterized by its negative feedback and OC and CC repetitions category is not strongly correlated with anything but with "I" and "like", making it a more subjective feedback category, where the tutor allows their opinion to show. It hence seems to be the case that engagement from the participant is greater in the presence of discussions marked with OC repetitions; when the tutor is likely providing feedback or backchannelling. Using ΔH turned especially useful when OC repetitions are not available, as they allowed to distinguish instruction turns from backchannelling, where engagement is greater. These results generalize the principles highlighted in [15] : an active initiator who wants an engaged audience should provide attentive listening and encourage participation.

V. CONCLUSION

Our observational study started with the hypothesis that engagement is dependent on the conversation unfolding in ways that can be perceived through the study of lexical spread. Looking at direct repetitions from one turn to another, but also the general spread of words in the conversation through entropy, we were able to observe that engagement is indeed dependent on the conversation unfolding through lexical spread. Engagement overall tends to be greater when the conversation's entropy decreases, and when people repeat each other. Moreover, roles in a conversation matter, and participants are more subtly engaging with the content of the tutors than the participants', as their engagement is then further correlated to differences in entropy and not – almost – solely to direct repetitions from the previous turn. More precisely, by exploring the content of the tutor's turns, it appears that minimal validations, backchannelling and providing feedback are correlated with greater engagement of the students in these online set-ups. The corpus upon which these observations are based was independently constructed, without bias from our expectations, and thus we think that results generalize to other data, but we are currently limited in knowing how far. Still, we have validated a method for (1) assessing engagement, which leads directly to suggestions for (2) fostering engagement. Future work may improve online meetings by incorporating some of these results into platform monitors and signals.

REFERENCES

[1] P. Baranyi, B. Berki, and Á. B. Csapó, "Concepts of cognitive infocommunications," *Infocommunications Journal*, pp. 37–48, 2024, Joint Special Issue on Cognitive Infocommunications and Cognitive Aspects of Virtual Reality, doi: 10.36244/ICJ.2024.5.5.

[2] A. Esposito, G. Cordasco, C. Vogel, and P. Baranyi, "Cognitive infocommunications," *Frontiers in Computer Science*, vol. 5, no. 1129898, pp. 4:1 129 898:1–4:1 129 898:7, 2023.

[3] P. Baranyi and A. Csapo, "Definition and synergies of cognitive infocommunications," *Acta Polytechnica Hungarica*, vol. 9, no. 1, pp. 67–83, 2012.

[4] ——, "Cognitive infocommunications: Coginfocom," in *2010 11th International Symposium on Computational Intelligence and Informatics (CINTI)*, 2010, pp. 141–146.

[5] C. Vogel and A. Esposito, "Advancing and validating models of cognitive architecture," in *10th IEEE Conference on Cognitive Infocommunications*, 2019, pp. 61–66.

[6] C. Vogel, M. Koutsombogera, and A. Esposito, "Aspects of Methodology for Interaction Analysis," in *2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Sep. 2020, pp. 000 141–000 146, iSSN: 2380-7350.

[7] C. Vogel and A. Esposito, "Interaction analysis and cognitive infocommunications," *Infocommunications Journal*, vol. 12, no. 1, pp. 2–9, 2020.

[8] P. Baranyi, A. Csapo, and G. Sallai, *CogInfoCom-Driven Research Areas*. Cham: Springer International Publishing, 2015, pp. 57–71. [Online]. Available: https://doi.org/10.1007/978-3-319-19608-4_5

[9] L. Kiss, B. P. Hámornik, D. Geszten, and K. Hercegfi, "The connection of the style of interactions and the collaboration in a virtual work environment," in *2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 2015, pp. 211–214.

[10] C. Navarretta, "Alignment of communicative behaviors and familiarity in first encounters," in *2014 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom)*, 2014, pp. 185–190.

[11] L. Hunyadi, "On multimodality in the perception of emotions from materials of the HuComTech corpus," in *2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 2015, pp. 489–492.

[12] J. Katona, "A review of human–computer interaction and virtual reality research fields in cognitive infocommunications," *Applied Sciences*, vol. 11, no. 6, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/6/2646

[13] I. Horváth, B. Berki, A. Sudár, Á. Csapó, and P. Baranyi, *Cognitive Aspects of Virtual Reality*. Cham, Switzerland: Springer, 2024.

[14] J. A. Fredricks, M. Filsecker, and M. A. Lawson, "Student engagement, context, and adjustment: Addressing definitional, measurement, and methodological issues," *Learning and Instruction*, vol. 43, pp. 1–4, Jun. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959475216300159

[15] A. Lane and M. L. Kent, "Dialogic Engagement," in *The Handbook of Communication Engagement*. John Wiley & Sons, Ltd, 2018, pp. 61–72. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119167600.ch5

[16] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School Engagement: Potential of the Concept, State of the Evidence," *Review of Educational Research*, vol. 74, no. 1, pp. 59–109, 2004, publisher: [Sage Publications, Inc., American Educational Research Association]. [Online]. Available: https://www.jstor.org/stable/3516061

[17] D. Garaialde, "Roomreader engagement measurement for breakout rooms," Apr 2023. [Online]. Available: osf.io/nxht2

[18] J. Reverdy, S. O'Connor Russell, L. Duquenne, D. Garaialde, B. Cowan, and N. Harte, "RoomReader: A Multimodal Corpus of Online Multiparty Conversational Interactions," in *LREC 2022*, Marseille, France, Jun. 2022, pp. 2517–2527.

[19] P. Sharma, S. Joshi, S. Gautam, S. Maharjan, S. R. Khanal, M. C. Reis, J. Barroso, and V. M. de Jesus Filipe, "Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning," in *Technology and Innovation in Learning, Teaching and Education*, A. Reis, J. Barroso, P. Martins, A. Jimoyiannis, R. Y.-M. Huang, and R. Henriques, Eds. Cham: Springer Nature Switzerland, 2022, pp. 52–68.

[20] K. J. Dickinson, K. E. Caldwell, E. A. Graviss, D. T. Nguyen, M. M. Awad, S. Tan, J. H. Winer, and K. Y. Pei, "Assessing learner engagement with virtual educational events: Development of the Virtual In-Class Engagement Measure (VIEM)," *The American Journal of Surgery*, vol. 222, no. 6, pp. 1044–1049, Dec. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S000296102100547X

[21] A. Rossi, M. Raiano, and S. Rossi, "Affective, Cognitive and Behavioural Engagement Detection for Human-robot Interaction in a Bartending Scenario," in *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, Aug. 2021, pp. 208–213, iSSN: 1944-9437. [Online]. Available: https://ieeexplore.ieee.org/document/9515435

[22] K. Suzuki, N. Hojo, K. Shinoda, S. Mizuno, and R. Masumura, "Participant-Pair-Wise Bottleneck Transformer for Engagement Estimation from Video Conversation," in *Interspeech 2024*. ISCA, Sep. 2024, pp. 4079–4083. [Online]. Available: https://www.isca-archive.org/interspeech_2024/suzuki24_interspeech.html

[23] R. Rafique, "Using Digital Tools to Enhance Student Engagement in Online Learning: An Action Research Study," in *Local Research and Glocal Perspectives in English Language Teaching: Teaching in Changing Times*, R. Khan, A. Bashir, B. L. Basu, and M. E. Uddin, Eds. Singapore: Springer Nature, 2022, pp. 229–248. [Online]. Available: https://doi.org/10.1007/978-981-19-6458-9_15

[24] M. Sharma, *Journal of Fundamental & Comparative Research*, vol. 8, no. 1, pp. 26–30, 2022.

[25] S. Wang, "Promoting Student's Online Engagement with Communication Tools," *Journal of Educational Technology Development and Exchange (JETDE)*, vol. 4, no. 1, Jun. 2011. [Online]. Available: https://aquila.usm.edu/jetde/vol4/iss1/8

[26] W. Hu, J. Tian, and Y. Li, "Enhancing student engagement in online collaborative writing through a generative AI-based conversational agent," *The Internet and Higher Education*, vol. 65, p. 100979, Apr. 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1096751624000411

[27] J. N. Bailenson, "Nonverbal Overload: A Theoretical Argument for the Causes of Zoom Fatigue," *Technology, Mind, and Behavior*, vol. 2, no. 1, Feb. 2021. [Online]. Available: https://tmb.apaopen.org/pub/nonverbal-overload/release/2

[28] G. Fauville, M. Luo, A. C. Muller Queiroz, J. N. Bailenson, and J. Hancock, "Nonverbal Mechanisms Predict Zoom Fatigue and Explain Why Women Experience Higher Levels than Men," Rochester, NY, Apr. 2021. [Online]. Available: https://papers.ssrn.com/abstract=3820035

[29] J. Martin, "Building Relationships and Increasing Engagement in the Virtual Classroom: Practical Tools for the Online Instructor," *Journal of Educators Online*, vol. 16, no. 1, Jan. 2019, eRIC Number: EJ1204379. [Online]. Available: https://eric.ed.gov/?id=EJ1204379

[30] Zoom Video Communications Inc., "Zoom meeting," 2019. [Online]. Available: https://zoom.us/meetings

[31] C. Oertel, G. Castellano, M. Chetouani, J. Nasir, M. Obaid, C. Pelachaud, and C. Peters, "Engagement in Human-Agent Interaction: An Overview," *Frontiers in Robotics and AI*, vol. 7, Aug. 2020, publisher: Frontiers. [Online]. Available: https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2020.00092/full

[32] D. Reitter and J. D. Moore, "Predicting success in dialogue," in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. Association for Computational Linguistics, 2007, pp. 808–815.

[33] D. Mekhaldi, "A study on multimodal document alignment: bridging the gap between textual documents and spoken language," Ph.D. dissertation, Faculty of Science, University of Fribourg (Switzerland), 2006.

[34] A. Mehler, A. Lücking, and P. Menke, "Assessing Lexical Alignment in Spontaneous Direction Dialogue Data by Means of a Lexicon Network Model," in *Computational Linguistics and Intelligent Text Processing*, A. F. Gelbukh, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, vol. 6608, pp. 368–379, series Title: Lecture Notes in Computer Science. [Online]. Available: http://link.springer.com/10.1007/978-3-642-19400-9_29

[35] J. Reverdy, M. Koutsombogera, and C. Vogel, "Linguistic Repetition in Three-Party Conversations," in *Neural Approaches to Dynamics of Signal Exchanges*, A. Esposito, M. Faundez-Zanuy, F. C. Morabito, and E. Pasero, Eds. Singapore: Springer Singapore, 2020, vol. 151, pp. 359–370, smart Innovation, Systems and Technologies. [Online]. Available: http://link.springer.com/10.1007/978-981-13-8950-4_32

[36] A. C. Murat, M. Koutsombogera, and C. Vogel, "Mutual Gaze and Linguistic Repetition in a Multimodal Corpus," in *Proceedings of the Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association (ELRA), 2022, pp. 2771–2780. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2022/pdf/2022.lrec-1.296.pdf

[37] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, Jul. 1948. [Online]. Available: https://ieeexplore.ieee.org/document/6773024

[38] M. Gnjatović, J. Tasevski, B. Borovac, and N. Maček, "An Entropy-Based Approach to Automatic Detection of Critical Changes in Human-Machine Interaction," in *2018 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Aug. 2018, pp. 000 175–000 178, iSSN: 2380-7350.