Measuring Synchrony in Dialog Transcripts

Carl Vogel and Lydia Behan

Computational Linguistics Group,
School of Computer Science and Statistics,
Trinity College, Dublin 2, Ireland
{vogel,behanl}@tcd.ie

Abstract. A finite register method of processing dialog transcripts is used to measure interlocutor synchrony. Successive contributions by participants are measured for word n-gram repetitions and temporal overlaps. The Zipfian distribution of words in language use leads to a natural expectation that random re-orderings of dialog contributions will unavoidably exhibit repetition – one might reasonably expect that the frequency of repetition in actual dialog is in fact best explained as a random effect. Accordingly, significance is assessed with respect to randomized contrast values. The contrasts are obtained from averages over randomized reorderings of dialog contributions with temporal spans of the revised dialogs guided by the original durations. Benchmark distributions for allo-repetition and self-repetition are established from existing dialog transcripts covering a pair of pragmatically different circumstances: ATR English language "lingua franca" discussions, Air-Traffic communications (Flight 1549 over the Hudson River). Repetition in actual dialog exceeds the frequency one might expect from a random process. Perhaps surprisingly from the perspective of using repetition as an index of synchrony, self-repetition significantly exceeds allo-repetition.

1 Background

Research into synchrony in dialog has deployed methods such as introducing delay, by using video to mediate communication, in a way that allows manipulation of whether interlocutors have access to partner contributions in real-time or with constructed delay. One type of study involves mother-infant communications mediated by video. The experimental paradigm makes the delay absolute, with a re-play condition seamlessly edited in between live interaction phases [15]. The striking effect is the disinterest expressed by the infant when the delay costs the illusion of interaction.

We focus here on two potential factors in the perception of interaction. It would be unsurprising if a politician revealed tactics for seeming engaged during meetings with the public, even when thinking about other matters entirely, as including occasionally repeating words or phrases uttered by their interlocutor or timing contributions to occasionally seem so interested in the content of the conversation to intervene through interruption or talk at the same time as interlocutors without actually taking the floor. In fact, repetition as an indication of participatory listening is just one of many functions of repetition in conversation [17]. Our inspection of interaction focuses on these measures. If these tactics are successful, then it must be because measures of lexical and sub-lexical repetition and of temporal overlap in natural conversations are significantly different than in un-natural conversation. Obviously, a conversation in which two interlocutors talk to each other about different topics would be quite un-natural, and any lexical overlap would be a matter of chance. Nonetheless, in natural dialog, it may be the case that overlaps and lexical repetitions are randomly distributed.

Even in reflecting on the fact that language is not a random process, considering that word rank-frequency distributions are Zipfian, one might reach the conclusion that repetitions are inevitable, and that the chances of repetition between dialog contributions are, in fact, best described by coin tosses. Such a position is motivated perhaps by the conviction that even if interlocutors are co-operatively talking about the same issue, they do so as independent agents, each making their own lexical decisions. Recent work has attempted to demonstrate the extent to which repetition in dialog correlates with task success [12, 13].

To explore these questions with a Monte Carlo approach,[1] from natural dialog transcripts we construct randomized versions of the turns. In what follow, “a turn”, “a contribution” and “a line” are synonymous expressions. They make sense when thinking of dialog transcriptions as scripts (within which actors have lines). Randomization of turns refers to re-ordering the turns with respect to each other, temporally, but not reordering within any turn. In the experiments reported here, we randomize real dialogs ten times. This means that the contributions of each participant are parsed into a data-structure in sequence from an actual transcript, and then each turn in the actual sequence is assigned a time-stamp in a range determined by the actual overall conversation duration. In the re-ordered dialogs, speakers still say the same thing overall, but not with any semblance of actual synchronization of contributions with respect to each other. Using this experimental framework of comparing measures in actual dialogs with counterpart measures averaged over randomized re-orderings, we focus on measures of lexical repetition by the speaker of others’ most recent contributions (allo-repetition), their own most recent contributions (self-repetition), and temporal overlap of contributions. With the focus in this methodology on repetition of components of most recent contributions, a fixed period is searched for potential repeated content. However, the actual temporal durations involved between a given contribution and the contributions which immediately precede it by each speaker may vary. The method we use contrasts with more powerful recurrence analysis techniques for identifying temporal coupling [14]; here, the window for anticipated repetition is structurally rather than durationally restricted.

In what follows, we first describe our register-based method of analyzing individual dialog contributions and durations with respect to the immediately prior turn of each speaker. We also describe the method of constructing randomized re-orderings of the turns by re-assigning pseudo-randomly determined

[1]The method fits into the “full-sample” dimension of the categorization provided by [11] with the labels “randomization”, “permutation” and “shuffling”.

start and stop times for each contribution. Within exemplars of several types of conversations, we assess levels of self-repetition and allo-repetition, towards an understanding of what counts as unmarked levels for both measures. We provide two case studies of analysis in this paradigm: §3.1 analyzes transcripts of dialogs involving five people over three sessions with a relevant feature that English provides a common ground; §3.4 examines a transcript of air-traffic communications, because extensive repetition is expected in this sort of dialog. We analyze the data with respect to levels of allo-repetition and self-repetition, and where possible, with respect to temporal overlap.

We find strong effects that separate actual dialog from randomized dialog: lexical expressions (from single words to sequences of up to five elements) are more likely to be repeated between contributions in actual dialog than in randomized dialog. Moreover, self-repetition effects are even stronger than repetition of others in ordinary chat. Overlap with others is also distinctive in actual dialog.

2 Methods

We analyze dialogs that have already been transcribed and are available on the web. Therefore, one issue of treatment that we do not have to address in this paper is the tokenization of the recorded dialog into tokens that might be deemed individual contributions of the speakers (1). At some point, one must make a decision between one contribution of Speaker $\mathcal{A}$ that has another speaker overlapping in the middle, and two separate contributions by Speaker $\mathcal{A}$. Making these decisions about the units of dialog constituents contributed by each speaker cannot be easy, and we do not revise any of the transcribers' decisions in this regard. We take a "line" of dialog to be an individual contribution of a speaker as attributed by a transcriptionist, the "lines" of a dialog is the partially ordered sequence of interleaved contributions. The transcripts are temporally ordered, but not totally so, given that contributions of interlocutors are interleaved. Each file is processed using a 'register' (3) for each speaker, initially empty, containing the contents of their most recent contribution (2).

(1) Ξ is the cast of actors (α) communicating.
(2) $u_j = \langle \tau_b, \tau_e, \alpha, \sigma \rangle$ is the j-th transcribed utterance:
τ_s, start time; τ_e, end time; α, actor; σ, statement
At u_j, $\alpha^{u_j} = \mathit{actor}(u_j)$; $\sigma^{u_j} = \mathit{statement}(u_j)$; etc.
(3) $\mathcal{R}_\alpha$, for each $\alpha \in \Xi$, is a register that records the start time, stop time and content of the last utterance of α.
$\mathcal{R}_\Xi$ refers to the set of registers;
$\mathcal{R}_{\Xi/\alpha}$ refers to the set of registers for all actors but α;
($g(n, \sigma)[i]$ denotes the i-th element of $g(n, \sigma)$).

Ultimately, for each utterance, count tokens shared with immediately preceding turns (their own (6), and their interlocutors' (5), in both cases, as given by (4) for the each length of $n-$gram to be counted) as recorded in the interlocutors' registers. The actual repetition values are then compared with those derived from some number (ten in the experiments reported here) of randomized re-orderings of the turns (AKA contributions). The constituent sequence of words within any individual contribution are left intact in their original order.

(4) $\kappa(n, \sigma^1, \sigma^2) = \sum_{i=1}^{g(n,\sigma^1)[max]}(g(n, \sigma^1)[i] \in g(n, \sigma^2))$
(5)

$$allo\text{-}shared(u_j, \mathcal{R}_{\Xi}, n) = \sum_{\alpha}^{\Xi/\alpha^{u_j}} \kappa(n, \sigma^{u_j}, \sigma^{\mathcal{R}_\alpha})$$

(6) a. first count with respect to former value for self:

$$self\text{-}shared(u_j, \mathcal{R}_{\Xi}, n) = \kappa(n, \sigma^{u_j}, \sigma^{\mathcal{R}_{\alpha^{u_j}}})$$

b. update register for the agent's own current utterance:

$$\mathcal{R}_{\alpha^{u_j}} := \langle \tau_s^{u_j}, \tau_e^{u_j}, \sigma^{u_j} \rangle$$

The random-reordering of the dialogs is effected by generating new start-times and durations for each utterance, and then sorting the utterances on their temporal indices. The times are selected using using random generators based on parameters that depend on the values in the original conversation (7). Thus, for each utterance u_i a re-indexing u'_i is constructed (8).

(7) given $u_1 \ldots u_{max}$
  a. $starttime = \tau_s^{u_1}$
  b. $stoptime = \tau_e^{u_{max}}$
  c. $maxoverlap$ = maximum temporal overlap of u_i and $\mathcal{R}_{\Xi}$ noted at time of shared n-gram computation.
(8) for each u_i,
  a. $\tau_s = rand(0, stoptime)$
  b. $\tau_e = \tau_s + rand(0, maxoverlap)$
  c. $u'_i = \langle \tau_s, \tau_e, \alpha^{u_i}, \sigma^{u_i} \rangle$

The u' are sorted on their value for τ_s. In the re-ordered dialog, we measure overlap, allo-shared tokens and self-shared tokens as before. Again, we consider n-gram sequences up to $n = 5$, and contrast "reality" with 10 randomizations.

Thus, an output file is generated which contains a temporal sequence of lines, each annotated for exactly one speaker and with appropriate measures for each line. One such line is constructed for each level of N-grams up to five. Repetitions of N-grams are recorded as counts with respect to the values in the registers as either SelfShared or OtherShared tokens. Here we focus on these as count values as opposed to ratios that relativize figures to the total number of N-grams that would have been possible to share between a dialog contribution and preceding contributions stored in registers.[2] Durations in seconds of temporal

[2]Given that we consider tokenization up to $N = 5$, we wanted to simplify treatment of utterances with fewer than N tokens, and thus avoid domain errors from division by zero.

overlaps with the most recent contributions of each speaker are also recorded for each line, as well as the count of the number of overlaps given the timestamps in each of the registers at the point of evaluation.

The annotations of times on the randomized dialogs interact with the computation of potential overlap in the DialogType=RANDOMIZED conditions. A start-time for each utterance is selected between 0 and the total number of seconds in the day's actual conversations. A corresponding stop-time for each utterance is recorded as the utterance's start-time plus an offset given by a random number of seconds between 0 and 10, the maximum overlap duration within the first dialog's actual conversations. Each shuffled dialog is determined by the old contributions sequenced by their new temporal order.

3 Application of Methods

3.1 Case Study 1: English as *Lingua Franca* in balanced chat

The data used here is that described by Campbell [5], conversations in English over three days inclusive of 5 speakers (two native-English speaking), at ATR in Japan.[3] The conversations are relatively balanced in the contributions made by participants, and when watches the accompanying video, one notices a high level of mutual engagement. It is reasonable to take these conversations as representative of engaged, balanced conversation. One speaker, g, was present only for the second day. We treated the data solely by regularizing the time-stamps to an HH:MM:SS format. We extracted columns of the data corresponding to timestamp, speaker id, and transcribed speech. We did not alter the transcriptions: editing errors are not addressed, nor are records of transcription difficulty ("@w") updated. The former would regularize spelling, and would increase the chances of repetition in all conditions. The latter would diminish current instances of repetition, separating the individual instances of the marker into more clear words. However, an interesting feature of the current edition of the transcripts is that the marker encodes unintelligibility of contributions. Whether speaker decision about utterance effort resulting in unintelligibility is conscious in these instances is a matter for debate; however, general effects of speaker intelligibility during the course of dialog are known (eg. Bard et al. [2]). There are 29900 lines of real dialog, and another 299000 from randomizations.

3.2 Results

Shared Expressions A binary variable, DialogType, records whether the measurements for an item correspond to a dialog contribution in its actual order or in a random one. In analyzing the data, the four levels of N greater than one were coalesced into a single level ("2+") of a related variable N'. Using a generalized linear model with a quasi-poisson error distribution, we separately

[3] We are grateful to Nick Campbell for use of the data from `www.speech-data.jp` – last verified February 2012.

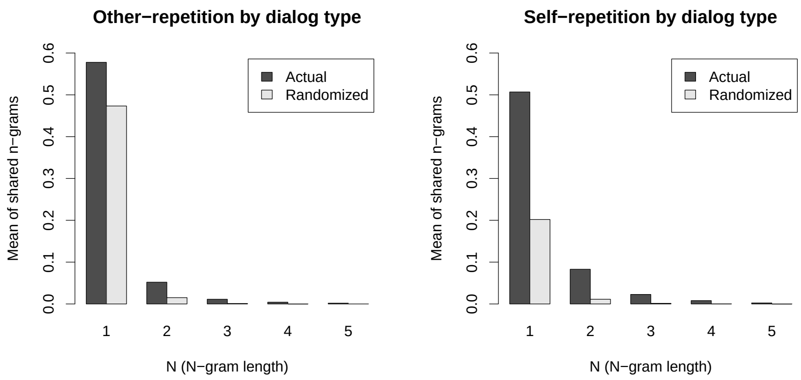

Fig. 1. *Actual* dialog vs *Randomized* turns: sharing others' (left) & own N-grams (right), by N

considered all interactions of DialogType × DAY × Speaker × N' on OtherShared and then on SelfShared. Figure 1 (L) graphs the distribution of mean scores for the count of OtherShared, and Fig. 1 (R) shows the same for SelfShared, both for each value of N. Significantly higher values for each value of N obtain in the actual dialogs than in the randomized dialogs for both repetition of others and of self. The effect of DialogType being set to actual in contrast to the randomized contrast is significantly higher values of OtherShared ($p < 0.005$) and of SelfShared ($p < 2 * 10^{-16}$). Interactions that do not include the factor DialogType=actual are not of interest: effects that obtain or which are commented upon as not emerging include the interaction with DialogType=actual. With respect to repetition of sequences in the preceding contributions of the others, the four-way interactions are not significant, nor the three way interactions. The actual orderings combined with $N' = 2+$ have the effect of significantly higher counts of OtherShared ($p < 3.1 * 10^{-5}$). No effects of DAY or Speaker emerged. Considering self-repetitions, there were significant positive effects of Speaker for g ($p < 0.02$) and $N' = 2+$ ($p < 1.3 * 10^{-9}$). There was a positive interaction for Speaker g with $N' = 2+$ ($p < 0.009$), and negative interaction for Speakers k and y with $N' = 2+$ ($p < 0.05$).

Figure 2 (L) shows the means of repetition of others by speakers, and (on the right) the means of self-repetitions. Self-repetition is systematically greater than repetition of others in the difference from the random values. Figure 3 shows continuity of the main effects over the three days. Interactions of effects are shown in Figures 4 and 5. In actual dialogs, the mean of repetitions for each day, speaker or choice of N in N-gram counts is at least the level in the randomized dialogs, or at a higher level. This holds for both allo-repetition and self-repetition, but the self-repetition values yield greater differences to the randomized values.

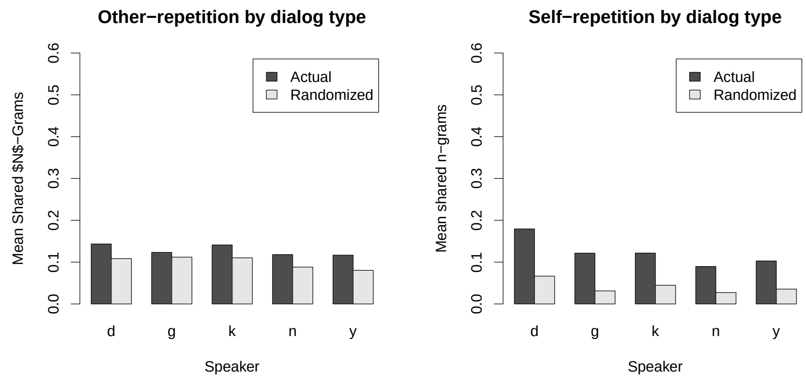

Fig. 2. Repetition in *Actual* vs *Randomized* dialog, of: others (L) & self (R), by Speaker

Temporal Overlap The approach to simulated overlap adopted in here is naive in allowing the contribution durations in seconds to be a random number between 0 and the longest overlap time. While this might seem to install a proclivity away from overlaps, it actually does not, as is illustrated in Figure 6. In any case, the distribution of actual overlaps is more sharply skewed than the overlaps in the randomized data. The actual dialogs show significantly less overlap than the randomized ($p < 2 * 10^{-16}$) dialogs. SPEAKER n exhibits significantly less overlap $p < 0.001$ than the randomized controls.[4] (The lesser amount of overlap involving speaker k approached significance.) Increased overlap was exhibited on Day 2 ($p < 2 * 10^{-16}$) and decreased overlap on Day 3 ($p < 1.5 * 10^{-6}$). An interaction with Day 2 and SPEAKERS k and n involves greater overlap for both ($p < 0.001$), and Day 3 for SPEAKER n ($p < 0.001$). No other factors studied have significant interactions jointly or in isolation with the condition where DIALOGTYPE=ACTUAL. That the actual data diverges so sharply from the random data may be an artifact of the particular simulation strategy used. Obviously, all of the effects reported are artifacts of the simulation strategy; however, alternative methods of assigning random temporality merit exploration.

3.3 Discussion

The direction of difference in actual dialogs between self-repetition and repetition of other interlocutors in the results is perhaps surprising. An analysis that was thus not anticipated at the outset reveals that the difference is significant. While the results reported show that the means allo-repetition and self-repetition are close, the difference between the means for allo-repetition and the randomized counterpart is smaller than the difference between actual self-repetition and its randomized counterpart. To quantify this, we constructed a

[4] `glm(OverlapSeconds~DialogType*Speaker*Day)`, quasipoisson error family.

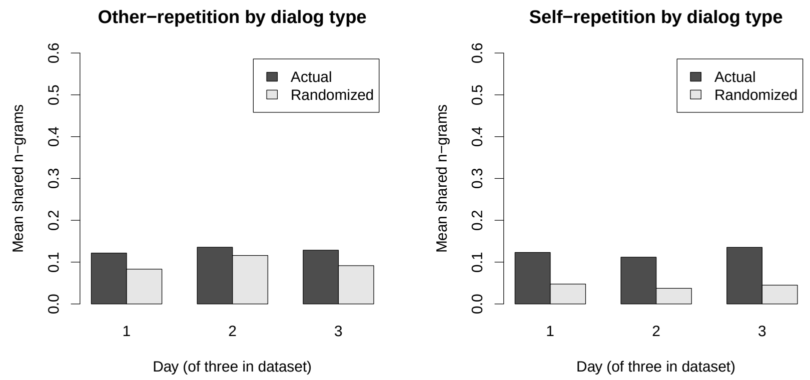

Fig. 3. Repetition in *Actual* vs *Randomized* dialog, of: others (L) & self (R), by Day

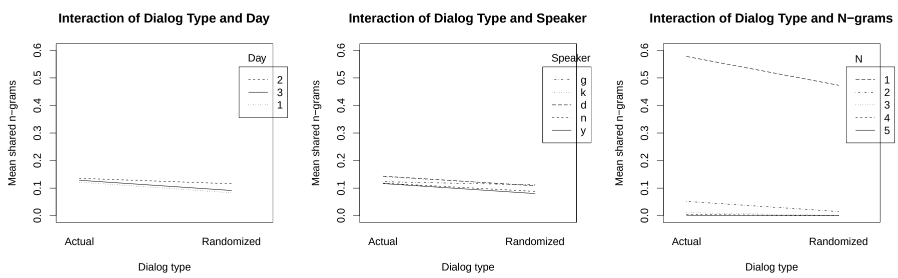

Fig. 4. Interactions of repetitions of others' N-grams: Dialog-Type vs Day (left), Speaker (middle), N (right)

variable DShared as SelfShared - OtherShared. Our reasoning was that if our perception that self-repetition is stronger than other-repetition, then the effects should be visible in this constructed variable. We reason that if the difference between SelfShared and OtherShared in the actual conversation is positive and significantly bigger than the randomized counterpart, then we have captured a difference that separates SelfShared from its randomized version as greater than OtherShared and its random counterpart. If the DShared value is negative, and the difference between the real and random version is significant, then it is the allo-repetition value that provides the greater difference (and significantly so). Our proxy measure of the relationship we are actually interested in is not the only possible one available to evaluate. We then tested effects on this variable using a generalized linear model with a Gaussian error

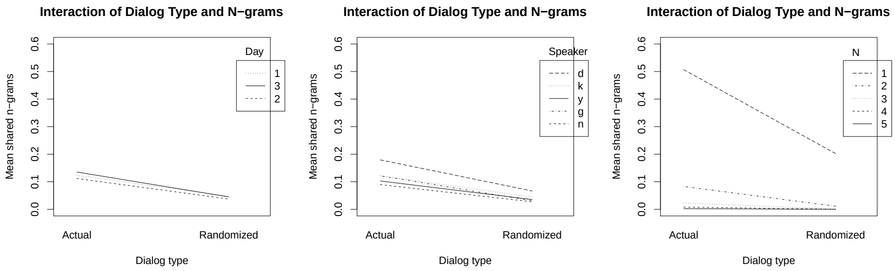

Fig. 5. Interactions of self-repetitions: Reality vs Day (left), Speaker (middle), N (right)

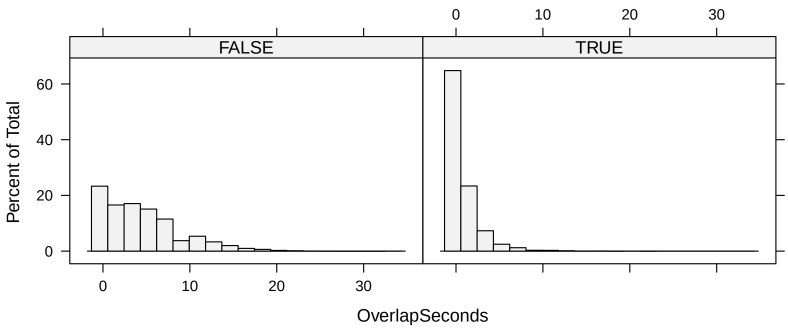

Fig. 6. Dialog Type *Actual* vs. *randomized* Mean Temporal Overlap Histograms

family.[5] There is, in fact, a significant positive effect of DialogType=actual on this variable ($p < 2 * 10^{-16}$).

3.4 Case Study 2: a crisis situation with a pre-defined leader

Given assumptions about repetition in air traffic communications (extensive) and overlaps (scarce), we decided to analyze a dialog from this setting. It happens that transcripts where incidents are involved are most readily available. We decided to examine the one recorded during the landing of US Airways Flight 1549 on the Hudson River on January 15, 2009. The transcript was prepared from the cockpit voice recorder [4]. Following consultation with a licensed flight instructor, we ignored all of the automated contributions; one of those, that of the Automated Terminal Information System (ATIS), is on a loop and the pilot must repeat "papa" when the relevant information is registered. There are 2860 dialog contributions in the resulting corpus. Thus, we consider the contributions

[5] `glm(DShared~DialogType)`

of 18 of the recorded channels: CAM, CAM-?, CAM-1, CAM-2, CLC, DEP, GND, HOT-?, HOT-1, HOT-2, INTR-1, INTR-4, PA-1, PA-2, RDO-1, RDO-2, RMP, TWR. Here, we ignore the fact that individuals choose different channels for different purposes (e.g. PA-1 and INTR-1 are both the captain, speaking to the passengers in one case and to ground crew in the other). In the analysis reported below, Voices are therefore individuated as the distinct sources: RDO-1, CAM-1, PA-1, HOT-1 and INTR-1 all contain the voice of the captain; RDO-2, CAM-2, PA-2, HOT-2 and INTR-2 all contain the voice of the first officer; the other channels are all analyzed as "AllElse".

Results Fig. 7 shows the mean sharing of N-grams between the ACTUAL and RANDOMIZED dialog types for allo-repetition and self-repetition. It can be seen that in this data set, with voices individuated in this way, there is more allo-repetition than self-repetition for each voice, but that the level is not uniformly greater in the actual dialog than in the randomized dialogs. Only the captain's voice displays a clear difference on both measures in this visualization. The univariate effect sizes from voice and dialog type on mean repetition is shown in Fig. 8: the captain and first officer show less repetition than the other voices recorded (the others individuated as one voice, in this analysis), and the effect of dialog type, with more repetition in actual dialog than in the randomized counterpart, being smaller. The interaction between voice and dialog type for allo-repetition and self-repetition is shown in Fig. 9. The effects of interest in allo-repetition were not significant. In the case of self-repetition, there is significance in the interaction, with the voices of both the captain ($p < 0.05$) and first officer ($p < 0.02$) providing more self-repetition in actual dialog than in the randomizations.[6] The greater difference in repetition between ACTUAL and RANDOMIZED dialogs as measured for self-repetition than for allo-repetition that appears in many other dialog contexts does not exist here.

The temporal overlap effects are shown in Figure 12 (overall (L), analyzing in terms of two distinct speakers vs. all else (M) and by individual channel (R)). Effects depending on DIALOGTYPE=ACTUAL were significant in interaction with the speaker: there was significantly more overlap with other speaking agents in the ACTUAL dialogs for both the captain and first officer than for participants generally. In Figure 12 (R), noting that HOT-1 is a channel used by the captain and HOT-2, by the co-pilot, it is clear that propensity to overlap temporally is not even among all participants in this dialog setting. Figure 10 demonstrates the tendency here is for more pronounced allo-repetition than self-repetition for each level of N (although the value of N is not significant as a univariate feature). Figure 11 (left) shows that this varies greatly with each speaker.

Discussion While repetition is consciously part of the system of air-traffic communication [6], the terseness that is also part of the ritual eliminates other aspects of the language which would ordinarily be open for unconscious repetition (e.g. there is an evident reluctance to use the preposition "to" in discussion

[6] Using `glm(SelfShared~DialogType*Voice)` and a quasipoisson error family.

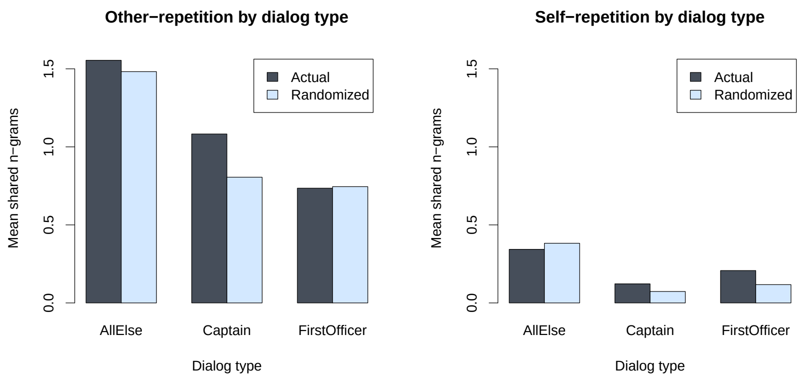

Fig. 7. Flight 1549 Speaker: Random vs Real allo-sharing (L) self-sharing (R)

of transit towards particular altitudes, lest it be confused with a numeral). By construction, air traffic communication in an emergency situation is not representative of air traffic communication in general. It is an open question how repetition is manifest during air-traffic communications outside crisis events. We have also analyzed this particular corpus with each channel individuated separately, including the separating the voices of the captain and first officer according to intended audience, and point out that strong effects associated with individual channels remain evident. What can be observed is that the pattern of repetition overall, and allo-repetition and self-repetition in isolation, vary with the participant. A generalization supported by this observation is that the participant's role matters as much as the individual filling the role.

4 General Discussion & Final Remarks

For final comparative discussion, we present graphs of the overall sharing of total n-grams for both of the case-studies discussed here in Figures (13)-(14). In both figures, the graph on the left indicates the amount of allo-sharing and the graph on the right shows self-sharing. Within each graph, the bar on the left indicates the levels of sharing in the dialogs as actually ordered, and on the bar on the right indicates the level of sharing in the randomizations. Figure 13 thus shows that in the ATR data, representative of dialogs in which partners show high levels of engagement and mutual interest, there is more self-sharing than allo-sharing, but higher levels of shared tokens in the real data than in the random data on both measures. The Flight 1549 cockpit conversations present rather more allo-repetition is evident than self-repetition. This is consistent with naive expectations of air-flight communications involving much repetition of others to signal understanding.

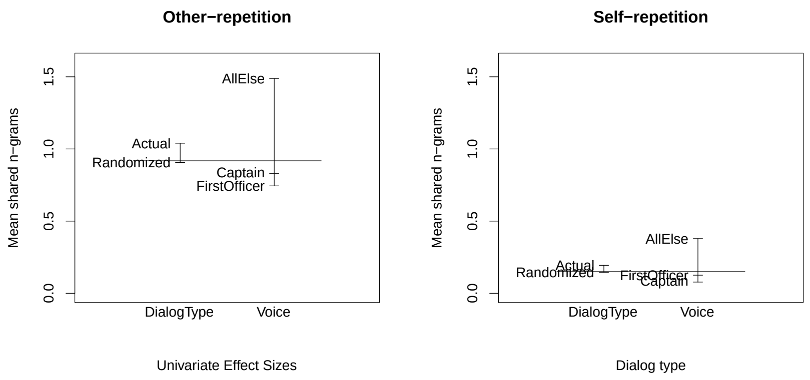

Fig. 8. Flight 1549 Speaker: Random vs Real allo-sharing (L) self-sharing (R)

Of course, it would be useful to explore these dialogs individually in greater depth, rather than dealing with them all in aggregate. We have also begun analysis of the MapTask dialogs using our methods as well [2] and the SwitchBoard dialogs [7, 8]. One of our aims is to use the method of quantifying interaction via repetition analysis to be able to assess the extent to which it is possible to reliably quantify the degree of synchrony that exists in a community in relation to the origins of synchronization: in some part, it arises through actual interactions; and in other parts, it emerges from agents independently articulating similar points and with the same linguistic expressions. We wish to characterize dialog types and participant role types in relation to patterns of quantified allo-repetition and self-repetition that are evident in the interactions. This has potential application in evaluating systems [16].

One might object to our methods arguing that it is wholly un-natural to consider transcripts instead of underlying audio recordings, or better, full-multimodal corpora such as Campbell [5] collected, or more fruitful to ignore text [3]. Researchers have demonstrated that very rich data sources can be tapped to measure interlocutor involvement in conversation, measuring articulation rates, voice intensity, etc. [10]. While applauding that work, part of our response is that if effects of synchrony can be detected even in the relatively impoverished record of textual transcripts (or interactions that might be recorded in text-based online communities), then it is important exploit this and to refine the means of detection and to establish an interpretative framework for understanding different levels of synchrony. Success implies that the resulting techniques of computational linguistics can contribute to assessment of the naturalness of patently fabricated scripts [9]. Thus, it is necessary further to examine additional dialogs in the light of such an analysis. A different objection is that the measures of temporal overlap used here to contrast with actual overlap are contrived. We would argue that the method is reasonable, but agree that there is more to explore here,

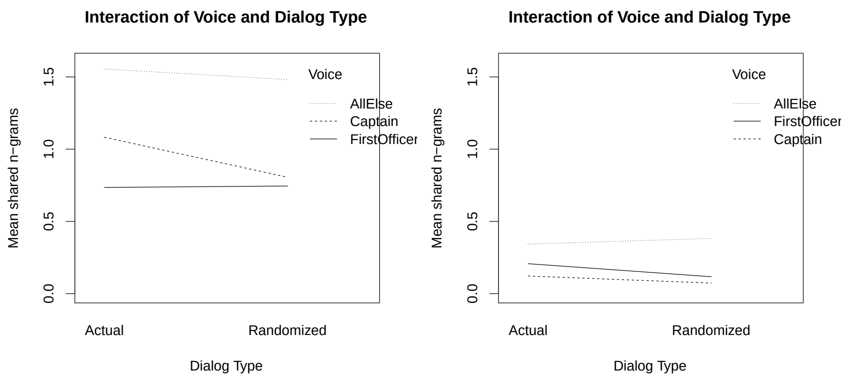

Fig. 9. Flight 1549 Speaker: Random vs Real allo-sharing (L) self-sharing (R)

including the more sophisticated and quite successful methods recently used by Altmann in analyzing synchronized body motions [1].

The dialogs analyzed may be treated as arbitrary in that they were not recorded with the analysis reported here in mind. Nonetheless, the effects reported in terms of greater overall repetition than with respect to random dialogs, and the effect of greater self-repetition than repetition of others may be integral to the particular data at hand. We think that divergences from this pattern have functional explanations (e.g, a direction giver repeats phrases less than chance would suggest, and personality types of participants matter). Nonetheless, gender, age, educational experience, and all of the other attributes of interlocutors that one might imagine interacting are all left unanalyzed at present.

We currently feel that an overall repetition effect, and more pronounced levels of self-repetition than allo-repetition constitute a signature of synchrony in natural dialog. The self-repetition preponderance (even at $N' = 2+$) may be partly explained by continued maintenance of a dialog plan and partly by the general effects of individual differences in language use that make authorship attribution viable.

Acknowledgements

We are grateful for feedback from anonymous reviewers of an earlier draft of this work for encouragement from Anna Esposito, leader of EU COST Action 2102.

References

1. Altmann, U.: Investigation of movement synchrony using windowed cross-lagged regression. In: Esposito, A., Vinciarelli, A., Vicsi, K., Pelauchaud, C., Nijholt,

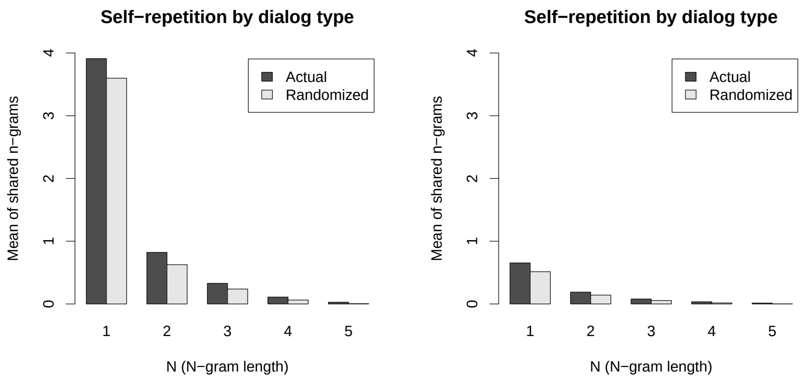

Fig. 10. Dialog type *ACTUAL* vs *RANDOM*: allo-sharing (L) & self-sharing of *N*-grams (R), by N

A. (eds.) Analysis of Verbal and Nonverbal Communication and Enactment: The Processing Issues, pp. 334–345. Heidelberg: Springer (2011)

2. Bard, E.G., Anderson, A., Sotillo, C., Aylett, M., Doherty-Sneddon, G., Newlands, A.: Controlling the intelligibility of referring expressions in dialogue. Journal of Memory and Language 42, 1–22 (2000)
3. Bouamrane, M.M., Luz, S.: An analytical evaluation of search by content and interaction patterns on multimodal meeting records. Multimedia Systems 13, 89–102 (2007), DOI10.1007/s00530-007-0087-8
4. Brazy, D.: Group chairman's factual report of investigation: Cockpit voice recorder dca09ma026 (2009), docket Number SA-532, Exhibit 12, National Transporation Safety Board `http://ntsb.gov/dockets/aviation/dca09ma026/420526.pdf`–last verified 06/2010
5. Campbell, N.: An audio-visual approach to measuring discourse synchrony in multimodal conversation data. In: Proceedings of Interspeech 2009 (2009)
6. Cushing, S.: Fatal Words: Communication Clashes and Aircraft Crashes. University of Chicago Press (1994)
7. Godfrey, J.J., Holliman, E.: Switchboard-1 release 2 (1997), linguistic Data Consortium
8. Graff, D., Bird, S.: Many uses, many annotations for large speech corpora: Switchboard and TDT as case studies. In: Proceedings of the Second International Conference on Langu age Resources and Evaluation. pp. 427–433. European Language Resources Association, Paris (2000)
9. Murtagh, F., Ganz, A., McKie, S.: The structure of narrative: the case of film scripts. CoRR (2008), http://arxiv.org/abs/0805.3799
10. Oertel, C., Looze, C.D., Scherer, S., Windmann, A., Wagner, P., Campbell, N.: Towards the automatic detection of involvement in conversation. In: Esposito, A., Vinciarelli, A., Vicsi, K., Pelauchaud, C., Nijholt, A. (eds.) Analysis of Verbal and Nonverbal Communication and Enactment: The Processing Issues, pp. 147–154. Heidelberg: Springer (2011)

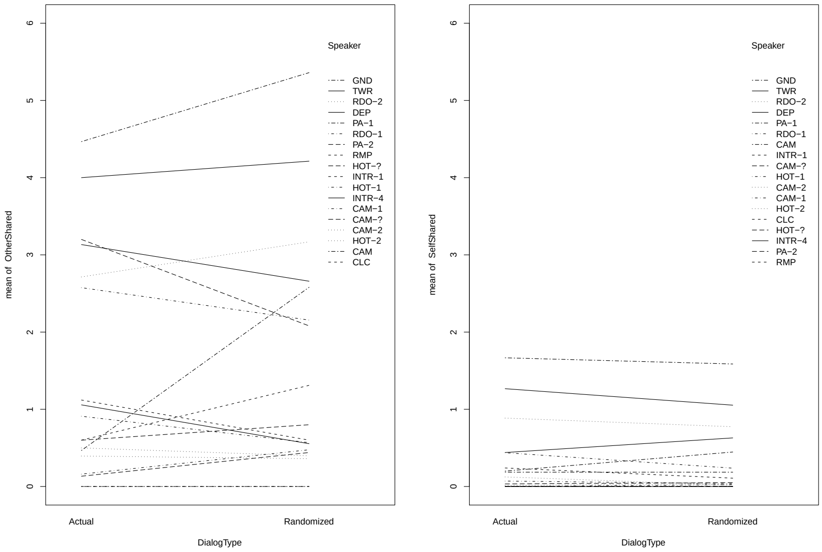

Fig. 11. Interaction of DialogType & Speaker: allo-repetition (left), self-repetition (right)

11. Ramseyer, F., Tschacher, W.: Nonverbal synchrony or random coincidence? how to tell the difference. In: Esposito, A., Campbell, N., Vogel, C., Hussain, A., Nijholt, A. (eds.) Development of Multimodal Interfaces: Active Listening and Synchrony, pp. 182–196. Springer, LNCS 5967 (2010)
12. Reitter, D., Keller, F., Moore, J.: Computational modeling of structural priming in dialogue. In: Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL. p. 121124. Association for Computational Linguistics (2006)
13. Reitter, D., Moore, J.: Predicting success in dialogue. In: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics. pp. 808–815. Association for Computational Linguistics (2007)
14. Richardson, D.C., Dale, R., Schokley, K.: Synchrony and swing in conversation: coordination, temporal dynamics and communication. In: Wachsmuth, I., Lenzen, M., Knoblich, G. (eds.) Embodied Communication. Oxford University Press (2008)
15. Stormark, K.M., Braarud, H.C.: Infants' sensitivity to social contingency: a "double video" study of face-to-face communication between 2- and 4-month-olds and their mothers. Infant Behavior & Development 27, 195–203 (2004)
16. Sun, X., Nijholt, A.: Multimodal embodied mimicry in interaction. In: Esposito, A., Vinciarelli, A., Vicsi, K., Pelauchaud, C., Nijholt, A. (eds.) Analysis of Verbal and Nonverbal Communication and Enactment: The Processing Issues, pp. 147–154. Heidelberg: Springer (2011)
17. Tannen, D.: Talking voices: repetition, dialogue, and imagery in conversational discourse. Cambridge University Press, Cambridge (2007)

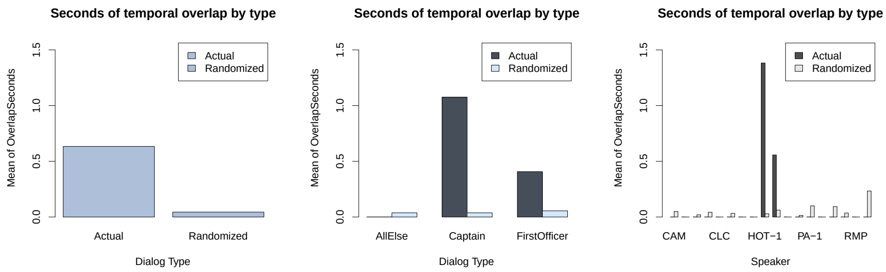

Fig. 12. Temporal overlap: Overall (L), by Speaker (M), by Channel (R)

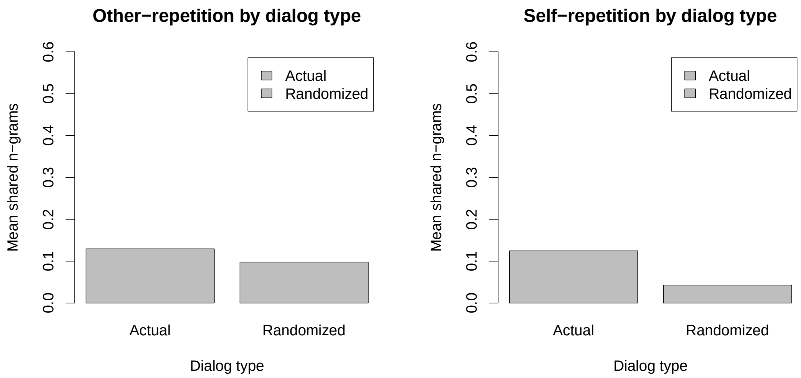

Fig. 13. ATR: Random vs Real allo-sharing (L) self-sharing (R)

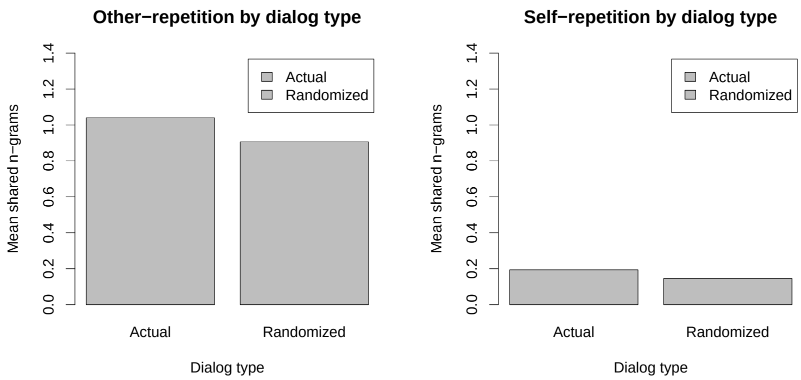

Fig. 14. Flight 1549 Speaker: Random vs Real allo-sharing (L) self-sharing (R)