

# Self-Repetition in Dialogue and Monologue

Patrick G. T. Healey and Matthew Purver

Queen Mary University of London,  
Cognitive Science Research Group,  
Mile End Road,  
London E1 4NS  
anon@authors.net

## Abstract

It has been claimed that natural dialogue is an especially repetitive form of language use. Comparison of dialogues and monologues in a corpus of naturally occurring speech (the DCPSE) suggests the reverse; monologue is substantially more repetitive than dialogue. We dub this the *bore* effect: the more people talk the more they repeat themselves. Dialogue, it appears, may provide an important means of escape from our cognitive and communicative ruts.

## 1 Repetition and Interaction

Work in psycholinguistics has sometimes characterised dialogue as an especially repetitive form of language use (Tannen, 2007; Pickering and Garrod, 2004; Pickering and Ferreira, 2008). However, previous research has indicated that, in free dialogue at least, repetition is rare. People repeat only 3% more of each other's words than would be expected by chance and systematically diverge from each other in their syntactic choices (Howes et al., 2010; Healey et al., 2014). This is compatible with a view of dialogue as constructive engagement in which participants respond to one another by actively building on, e.g.: modifying, adapting or elaborating each other's contributions rather than repeating them (Healey et al., 2014; Healey et al., 2018).

The principal evidence against repetition in natural conversation comes from the analysis of *other-repetition* (Howes et al., 2010; Healey et al., 2014). Spoken monologues, such as one-sided conversations or speeches, provide an interesting alternative test case that allows us to examine patterns of *self-repetition*. Do people repeat themselves more in monologues or dialogues? A constructive engagement view would predict that dialogue should **reduce** self-repetition, as people actively respond to each other's contributions. This contrasts with priming models that claim that repetition in dialogue is typically either **equivalent to or stronger than** in monologue (Pickering and Garrod, 2004; Pickering and Ferreira, 2008).

## 2 Method

The Diachronic Corpus of Present-Day Spoken English (DCPSE) includes samples ranging from face-to-face conversations to prepared speeches. The monologue collection used here was created by selecting all DCPSE files in which only one person spoke; this includes data from genres including radio broadcasts, sports commentary, sermons and lectures. The dialogue collection includes all dyadic conversations; this includes not only informal conversation but academic interviews, broadcast interviews and multi-party sports commentary. For the dialogue samples, we follow Healey et al. (2014), calculating lexical and syntactic similarity scores between each speaker turn and the preceding five turns by the same participant. For the monologue sample, the same calculations are made, but between sentences rather than speaker turns (the notion of speaker turn being irrelevant in monologue); we use sentence boundaries as annotated in the DCPSE. This produces 254 dialogue samples with an average of 45 turns and 736 words per speaker, and 106 monologue samples with an average of 74 sentences and 1097 words per speaker. Average turn length in the dialogues is 16.3 words, average sentence length in the monologues

14.7 words. Note that distances between dialogue speaker turns are greater than the distances between monologue sentences, because of the interleaving turns of the interlocutor.

The similarity calculation is based on the number of matches between candidate turns/sentences, using a standard kernel normalisation for length of sentence (see (Moschitti, 2006)):

$$N_{AB}/\sqrt{N_{AA} \times N_{BB}}$$

Here,  $N_{AB}$  represents the number of matching elements between turn/sentence  $A$  and turn/sentence  $B$  (words for lexical similarity; syntactic production rule subtrees for syntactic similarity), and  $N_{AA}$  the number of matches when  $A$  is matched against itself (see (Healey et al., 2014)).

### 3 Results

The basic pattern of results is illustrated in Figure 1 (the statistical analysis of these patterns are given below). The most obvious difference between the two graphs is that levels of syntactic repetition are higher than levels of lexical repetition. This is because there are substantially fewer possible syntactic constructions than there are possible lexical items. This difference is also reflected in the chance levels of repetition calculated by randomly re-ordering all of each person’s sentences/turns respectively and calculating the lexical and syntactic match in the same way as for the real samples. Chance repetition is higher for syntax (0.41 for monologue and 0.30 for dialogue) than for words (0.14 for monologue and 0.12 for dialogue).

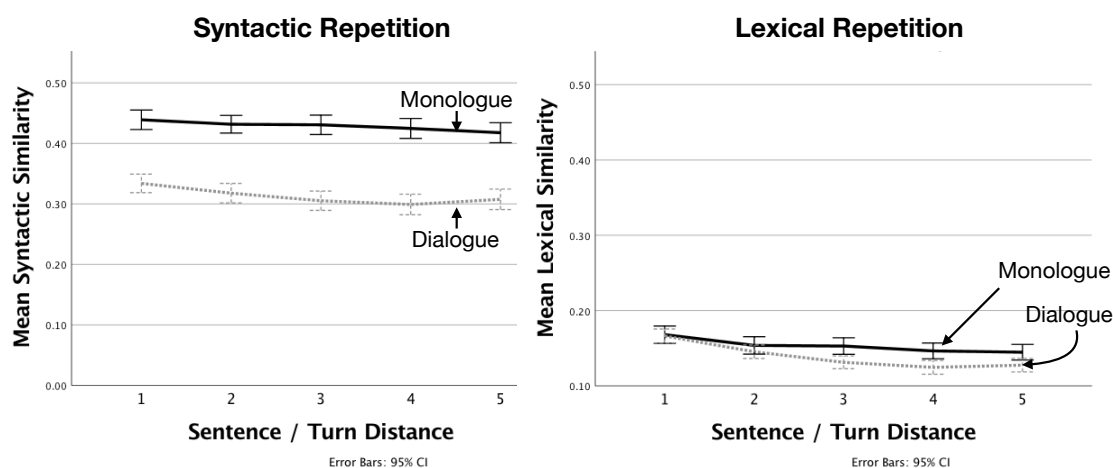


Figure 1: Patterns of Repetition Across Turns

Comparison of the patterns of self-repetition for monologue and dialogue indicates that there is more lexical repetition overall in monologue but this effect only reliably emerges at larger turn/sentence distances. In contrast to this syntactic repetition shows a more marked difference and is consistently higher in monologue at all sentence distances. This is highlighted by the observation that even after five intervening sentences people are still substantially more likely to repeat the syntax of their original sentence in monologue than they are after only one intervening turn in dialogue. In addition the graphs indicate a general tendency in both monologue and dialogue for likelihood of repetition (lexical or syntactic) to reduce with distance.

Two Generalized Linear Mixed Model (GLMM) analyses described below provide statistical tests of these effects. They also include a factor not captured in Figure 1: the amount people speak, measured here as total number of words produced. The GLMM analyses include Mode (Dialogue vs. Monologue), Distance (1-5 Sentences/Turns) and Words (total produced each speaker) are included as fixed factors, plus the Words  $\times$  Mode and Words  $\times$  Distance interactions, and Speaker as a random intercept.

**Lexical repetition** there is no simple main effect of Mode ( $F_{(1,1759)} = 3.65, p = 0.06$ ) and no Words  $\times$  Mode interaction ( $F_{(1,1759)} = 2.77, p < 0.09$ ) but there are main effects of Distance

( $F_{(1,1759)} = 42.6, p < 0.00$ ) and Words ( $F_{(1,1759)} = 42.6, p < 0.00$ ) and a Mode  $\times$  Distance interaction ( $F_{(1,1759)} = 4.95, p < 0.00$ ). The interaction shows that the difference in lexical repetition in dialogue is only statistically significant at distances of greater than 3 turns/sentences.

**Syntactic repetition** shows simple main effects of Mode ( $F_{(1,1759)} = 59.2, p < 0.00$ ), Distance ( $F_{(1,1759)} = 12.8, p < 0.00$ ) and Words ( $F_{(1,1759)} = 49.8, p < 0.00$ ); there are also Mode  $\times$  Distance ( $F_{(1,1759)} = 2.68, p = 0.03$ ) and Mode  $\times$  Words ( $F_{(1,1759)} = 8.16, p < 0.00$ ) interactions. The first interaction indicates that distance has a stronger effect on reducing syntactic repetition in dialogue. The second interaction indicates that the effect of talking more has a stronger effect on promoting repetition in dialogue.

### 3.1 Conclusion

Monologue, not dialogue, appears to be the more repetitive form of language use. The more people talk the more they repeat the words and syntax of their preceding turns. It seems natural to gloss this as the *bore* effect. The results seem clear but there are several possible explanations for them.

The effect might be due, in part, to genre: the monologue and dialogue collections here do not cover identical genres of conversation (although they both cover a range of genres, see above). There is an intuition that, for example, repetition for rhetorical effect might be an important characteristic of some forms of monologue such as sermons and lectures. Nonetheless, this doesn't account for the observation that as people talk more (total words) they are more likely to repeat themselves. This effect is found in *both* the monologue and dialogue samples and therefore does not appear to be explainable in terms of genre differences. Talking more in a conversation or speaking longer in a lecture both lead to significantly more repetition. Another simple possibility is that the delays between turns at speaking caused by other people's turns in dialogue cause decay or forgetting that leads to reduced repetition whereas in monologue there is no delay between successive turns. This does not easily explain the difference in syntactic repetition which remains marked at all distances including comparison of a turn distance of 1 with a sentence difference of five. More importantly, this explanation treats intervening turns as delays and ignores what they are doing as part of the dialogue.

Our interpretation is that monologues are more repetitive because without the stimulus of contributions from others we are more likely to slip into our habitual linguistic routines. Effective conversation depends on responding constructively to each other by building on what our conversational partners say and this helps to overcome our regressive tendency to bore.

## References

- Patrick G.T. Healey, Matthew Purver, and Christine Howes. 2014. Divergence in dialogue. *PLoS ONE*, 9(6):e98598, June.
- Patrick G.T. Healey, Gregory J Mills, Arash Eshghi, and Christine Howes. 2018. Running repairs: Coordinating meaning in dialogue. *Topics in Cognitive Science*, 10(2):367–388.
- Christine Howes, Patrick G. T. Healey, and Matthew Purver. 2010. Tracking lexical and syntactic alignment in conversation. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, Portland, OR.
- A. Moschitti. 2006. Making tree kernels practical for natural language learning. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*.
- Martin J. Pickering and Victor S. Ferreira. 2008. Structural priming: A critical review. *Psychological Bulletin*, 134(3):427–459.
- Martin Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27:169–226.
- D. Tannen. 2007. *Talking Voices: Repetition, Dialogue and Imagery in Conversational Discourse*. Cambridge: Cambridge University Press. Second Edition.