SPACER: A Parallel Dataset of *S*peech *P*roduction *A*nd *C*omprehension of *E*rror *R*epairs

Jiaxuan Li*, Shiva Upadhye* (* = equal contribution), Rutvik Gandhasri & Richard Futrell
University of California Irvine {jiaxul19, upadhyes, rgandhas,rfutrell}@uci.edu

Introduction Speech errors provide unique perspectives into the psychological mechanisms underlying online language processing. Speakers monitor their overt production and take corrective action when needed while comprehenders correct the perceived erroneous signal to infer the intended meaning. In the language production research, speech error patterns have been cited as evidence that a speaker's error-detection mechanism may be more sensitive to intercepting certain categories of errors [1, 2]. Studies focusing on language comprehension, on the other hand, have found that comprehenders exhibit a bias toward detecting and correcting errors with similar surface forms to the intended utterance [3, 4]. Whereas these findings provide crucial insights into the role of error processing in language use, existing datasets or experimental paradigms that investigate these mechanisms typically focus either on production or comprehension, making it difficult to compare the asymmetries and investigate the potential interaction between two systems. This study presents a parallel dataset of Speech Production and Comprehension Error Repairs (*SPACER*) that captures how naturalistic speech errors are corrected by speakers and comprehenders, thereby offering opportunities for comparative analyses across modalities and facilitating a comprehensive examination of the interplay between production and comprehension processes. The data collection is ongoing.

Methods Production: We focus on a particular kind of speech error known as *paraphasia* or lexical substitution. Specifically, we extract utterances with lexical substitutions and repairs from Switchboard NXT Annotations [5] – a disfluency annotated corpus of conversational speech [6]. An utterance was identified as having a substitution and repair if (i) the error (annotated as a *reparandum*) was immediately followed by a correction (annotated as a repair) (Example 1a.) or (ii) the speaker retraced or repeated previous material almost verbatim except for a single-word substitution (see Example 1b.). We include both these cases since they reflect a *fast* versus *slow* repair process respectively. We only consider utterances with an equal number of reparanda and repairs and therefore exclude instances of additions, deletions, contractions, and structural revisions. Utterances with multiple substitution errors were processed into frames with only a single error-repair pair (see Example 2). Additionally, disfluencies such as false starts, repetitions, or filled pauses were removed. We further manually eliminated items that require domain-specific world knowledge or include inappropriate discourse. We also introduce error-free filler that match the length of erroneous sentences. The dataset includes 576 instances of lexical substitutions or repairs and 480 instances of error-free utterances. See Fig. 1 for distribution of different types of error-repair pairs.

Comprehension We conducted human error correction experiments to understand how utterances are corrected during comprehension (N = 66). Each subject was presented with 48 critical utterances (Example 3, *italicized*) and 40 error-free fillers extracted from Switchboard. Comprehenders are instructed to check the quality of the last sentence from speech transcriptions and make necessary changes to recover the intended meaning of the speaker as fast as they can. Each selected sentence will be checked by at least ten comprehenders (Example 4), and annotated as *not corrected* and *corrected*.

Implications and Future Directions We presented a parallel dataset of naturalistic speech errors corrected by speakers and comprehenders. We plan to develop an integrated model [7] that predicts error correction behaviors by speakers and comprehenders. We predict that speakers and comprehenders might weigh linguistic cues differently during error correction, and their asymmetries might reflect iterative reasoning between comprehension and production.

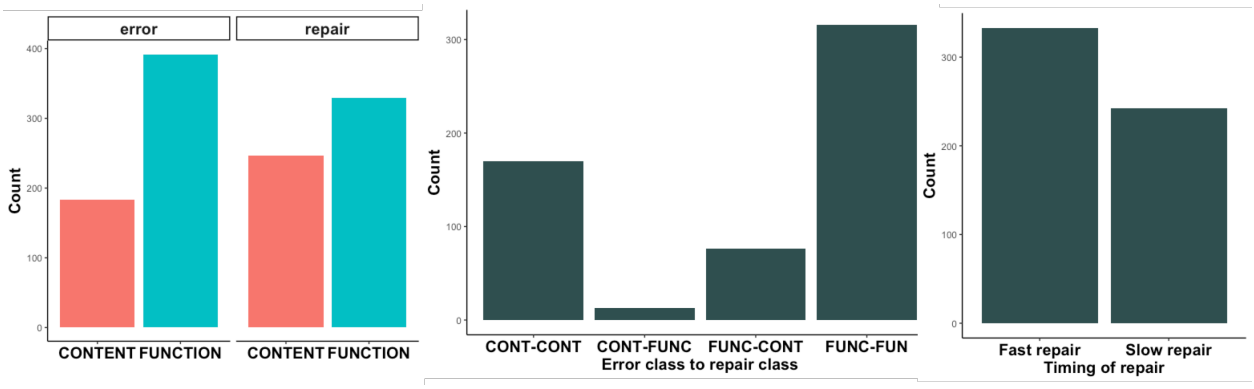

Figure 1: Left: Number of content and function word errors and repairs. Middle: Number of lexical substitutions where the error and repair belong to the same content (CONT) or function (FUNC) class (CONT-CONT and FUNC-FUNC) or different classes (CONT-FUNC and FUNC-CONT). Right: Number of fast and slow corrections

Example 1: Fast and slow repairs

1a. Fast Repair: I I think that might be *talking* **referring** to uh something kind of uh alternative to the draft you know

1b. Slow Repair: So until I see the entire quote old guard of the Soviet *military* of the Soviet **government** completely roll over and disappear preferably buried I still consider them a threat

Example 2: Pre-processing Pipeline

Original Utterance: But the *book* is the **magazine** is full of these mail-order *computer* mail order **houses**

2a. But the [book/magazine] is full of these mail order houses

2b. But the book is full of these mail order [computer/houses]

Example 3: Sentence presented to comprehenders

[Speaker A]: strangely because uh it's talking about a requirement for public service

[Speaker A]: and of course Pe- the Peace Corps the Boy Scouts is a is a voluntary activity service

[Speaker B]: Right

[Speaker A]: *I think that might be talking to something kind of alternative to the draft*

Example 4: Parallel correction

Original utterance: I think that might be *talking* to something...

Speaker correction: I think that might be **referring** to something...

Comprehender correction 1, 2 & 4: I think that might be **referring** to something

Comprehender correction 3: I think that might be talking **about** something

Comprehender correction 5: I think that might be *talking* to something...

Selected References [1] Levelt, W. J. M. (1989). Speaking: From intention to articulation. [2] Hartsuiker (2007). *Language and Cognitive Processes*. [3] Gibson et al. (2013). *Psychological science*. [4] Ryskin et al. (2018). *Cognition*. [5] Calhoun et al. (2010). *Language resources and evaluation*. [6] Godfrey et al. (1992). *IEEE*. [7] Upadhye, Li & Futrell (2023). *SCiL 2023*