[Proceedings of the Society for Computation in Linguistics](https://scholarworks.umass.edu/scil)

Volume 6 Article 40

6-1-2023

Bridging production and comprehension: Toward an integrated computational model of error correction

Shiva Upadhye
University of California, Irvine, upadhyes@uci.edu

Jiaxuan Li
University of California, Irvine, jiaxuan.li@uci.edu

Richard Futrell
University of California, Irvine, rfutrell@uci.edu

Recommended Citation

Upadhye, Shiva; Li, Jiaxuan; and Futrell, Richard (2023) "Bridging production and comprehension: Toward an integrated computational model of error correction," *Proceedings of the Society for Computation in Linguistics*: Vol. 6, Article 40.
DOI: https://doi.org/10.7275/q7mg-t564
Available at: https://scholarworks.umass.edu/scil/vol6/iss1/40

Bridging production and comprehension: Toward an integrated computational model of error correction

Shiva Upadhye*, Jiaxuan Li*, and **Richard Futrell**
{upadhyes,jiaxuan.li,rfutrell}@uci.edu
Department of Language Science
University of California, Irvine
*= Equal contribution

Abstract

Error correction in production and comprehension has traditionally been studied separately. In real-time communication, however, correction may not only depend on speaker or comprehender-internal preferences, but also the interlocutors' knowledge of each other's strategies. We present an integrated computational framework for error correction in both production and comprehension systems. Modeling error correction as Bayesian inference, we propose that both speaker and comprehender's correction strategies are influenced by their prior expectations about the intended message and their knowledge of a noise monitoring model. Our results indicate that speakers and comprehenders tend to weigh phonological and semantic cues differently, and these different strategies illuminate the asymmetries and potential interactions between the two systems.
Keywords: error correction; Bayesian inference; strategic cue-weighting;

Effective communication hinges on the ability to detect and correct inevitable errors in real-time: speakers monitor their overt speech and self-repair disfluencies on the fly while comprehenders make corrections to the perceived noisy signal to recover the speakers' intended message. Prior work has found that both speakers and comprehenders are sensitive to phonological and semantic features in error monitoring. Whereas comprehenders may be more likely to correct errors with a change of surface form (Gibson et al., 2013), a speaker's monitoring system may be differentially sensitive to combinations of semantic and phonological features (Hartsuiker, 2006). Intriguingly, most prior work on error monitoring and correction has focused exclusively on either speaker-internal or comprehender-internal error monitoring processes. In a communicative context, however, interlocutors also reason over the state of each other's knowledge and beliefs (Frank and Goodman, 2012; Goodman and Stuhlmüller, 2012). Consequently, speakers and listeners may also reason over noise in the communication channel (Bergen and Goodman, 2015), and adapt their usage to mitigate confusion (Buz et al., 2016). We develop a unified framework to investigate the strategic use of phonological and semantic information in the error correction process in both comprehension and production, and the potential interaction between the two systems.

We formalize error correction as a Bayesian rational inference about intended message. We learn weights for phonological and semantic factors in the model by fitting to naturalistic speech data for production and to offline reading and editing experiments for comprehension. The results indicate that speakers and comprehenders tend to use different strategies to perform error correction, which are related to inherent asymmetries between production and comprehension. Our model sheds light on strategic cue weighting in error monitoring, and bridges comprehension and production.

Model Fig. 1 shows an overview of the critical processes. We model the error correction process as Bayesian inference (Eq. 1):

$$p(x_i \mid x_p) \propto p(x_i)p(x_p \mid x_i), \tag{1}$$

where x_i is an intended message and x_p is a perceived message. We propose that the decision to repair is influenced by the prior knowledge of the intended message as well as a noise model that approximates how the utterance could be distorted by noise. We specify this noise monitor model as a strategic cue weighting of phonological and semantic distances (Eq. 2):

$$\Phi(x_p, x_i) = \underbrace{-[\alpha \mathrm{Phon}(x_p, x_i) + \beta \mathrm{Sem}(x_p, x_i)]}_{\text{noise monitor}}, \tag{2}$$

Proceedings of the Society for Computation in Linguistics (SCiL) 2023, pages 389-391.
Amherst, Massachusetts, June 15-17, 2023

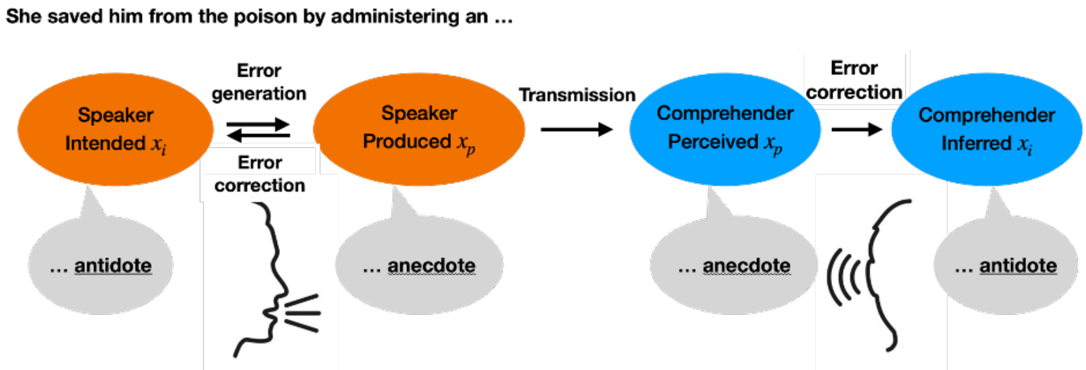

Figure 1: An overview of model architecture

where Phon is a phonological distance measure and Sem is a semantic distance measure. The likelihood term in Eq. 1, therefore, yields the probability that x_i will get distorted into x_p due to interference or perceptual noise (Eq. 3):

$$p(x_p \mid x_i) = \frac{e^{\Phi(x_p,x_i)}}{\sum_{x_i' \in C} e^{\Phi(x_p,x_i')}}. \quad (3)$$

where C is the set of possible repairs including the target x_i.

Data We fit the weights α and β in the noise model to production and comprehension data separately. For production, we extract utterances with single-word lexical substitution errors ($N = 1024$) from the Fromkin Speech Error database (Fromkin, 2000). The utterances are coded for whether or not the speaker chose to repair their error. For comprehension, we use the human sentence correction data from Ryskin et al. (2021). The dataset consists of 480 sentences with either semantic or syntactic errors at the end of the sentence. Each sentence is corrected by native English speakers in an offline sentence correction experiment, resulting in $N = 22,041$ corrections (Table 1).

	Example
Production	it's cold and CAMP there (**not corrected**) it is the question of the HOUR – of the fortnight (**corrected**)
Comprehension	She saved him from the poison by administering an *anecdote* **Corrections**: antidote, epipen, anecdote

Table 1: Example of stimuli for modeling production and comprehension-side correction processes

Implementation We estimate the prior probability of the intended message $P(x_i)$ by masking the target using XLNet (Yang et al., 2019). We approximate the noise model by computing the semantic and phonological distance between target x_i and error x_p using pre-trained GloVe embeddings and phonemic-feature based edit distance respectively. We fit the production model by choosing weights to minimize cross entropy loss in predicting ground truth (whether an error was corrected or not). For the comprehension model, we minimize the cross-entropy between model estimated and empirical probabilistic distribution over corrections.

Results Fig. 2 shows the phonological and semantic weights averaged over 1000 simulations. A positive weight indicates that a phonologically/semantically *similar* error is more likely to be corrected. We observe an asymmetry in the comprehender and speaker α: whereas comprehenders are more likely to repair phonologically similar errors, speakers show a show a mild preference toward correcting phonologically *dissimilar* errors. Furthermore, comprehenders and speakers also differ in the way they weight semantic information.

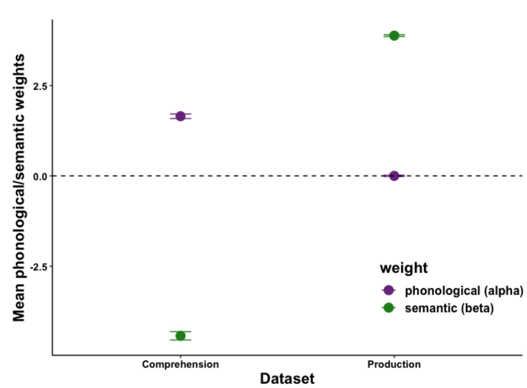

Figure 2: Averaged phonological and semantic weights for comprehension and production

Particularly, while comprehenders are more likely to correct semantically *dissimilar* errors as evinced by the negative β, speakers exhibit a preference toward correcting semantic competitors.

Discussion Our results indicate a key relationship between comprehension and production from asymmetries of error correction strategies.

We find that a comprehender is less likely to correct semantically compelling errors. We attribute this to the inaccessibility of the intended meaning to comprehenders. Given that a semantic competitor could be semantically appropriate (*She saved him from the poison by administering an* ***antidote/epipen***), comprehenders might not be able to detect the existence of a semantic errors. As a result, a speaker is incentivized to correct a semantic competitor. On the other hand, comprehenders tend to correct phonologically similar errors, phonological errors are still noticeable to comprehenders providing other contextual information (*She saved him from the poison by administering an* ***antidote/anecdote***). A speaker will therefore be less motivated to correct phonologically similar errors given these errors are recoverable for comprehenders. This reveals the potential interaction with the comprehension system: a speaker with knowledge of the comprehender's correction model is motivated to correct errors that a comprehender is incapable of recovering.

Some findings might be related to characteristics of the dataset selected. In comprehension dataset, the errors purposefully introduced at the end of the sentences occur at the end of the sentence and comprehenders were explicitly asked to perform error correction task, which might lead to a high error correction rate and a particular distribution of error types. The dataset of speech errors used for production, on the other hand, is likely to only include errors that may have been detected by the annotators. In the future, it is important to construct more parallel datasets in comprehension and production, and conduct more detailed analysis on the errors types in the dataset.

In conclusion, we observe an asymmetry in how speaker and comprehender strategies for correcting errors. We argue that the strategic use of phonological and semantic cues reflect potential interaction between comprehension and production. Speakers adopt different error correction strategies based on their understanding of the comprehenders' noise model, and comprehenders weigh phonological and semantic information differently to deal with information inaccessibility.

References

Leon Bergen and Noah D Goodman. 2015. The strategic use of noise in pragmatic reasoning. *Topics in cognitive science*, 7(2):336–350.

Esteban Buz, Michael K Tanenhaus, and T Florian Jaeger. 2016. Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of memory and language*, 89:68–86.

Michael C Frank and Noah D Goodman. 2012. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998.

Victoria A Fromkin. 2000. Fromkin's speech error database. *Online database, Max Planck Institute for Psycholinguistics, Nijmegen*.

Edward Gibson, Leon Bergen, and Steven T Piantadosi. 2013. Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, 110(20):8051–8056.

Noah D. Goodman and Andreas Stuhlmüller. 2012. Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science*, 5 1:173–84.

Robert J. Hartsuiker. 2006. Are speech error patterns affected by a monitoring bias? *Language and Cognitive Processes*, 21:856 – 891.

Rachel Ryskin, Laura Stearns, Leon Bergen, Marianna Eddy, Evelina Fedorenko, and Edward Gibson. 2021. An erp index of real-time error correction within a noisy-channel framework of human communication. *Neuropsychologia*, 158:107855.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32.