Crosslinguistic Corpus of Hesitation Phenomena: A corpus for investigating first and second language speech performance

Ralph L. Rose <rose@waseda.jp> Waseda University Faculty of Science and Engineering, Tokyo Japan

Abstract

A consensus is forming that there is a need to evaluate second language (L2) speech performance with respect to first language speech behavior. To support this need, the Crosslinguistic Corpus of Hesitation Phenomena was developed. This freely available corpus is designed to investigate the crosslinguistic influence of speech patterns and consists of recordings of speakers producing first and second language speech samples in response to parallel elicitation tasks in each language. Preliminary results from the corpus are consistent with other findings that second language performance is sometimes correlated with first language behavior [3,7]. In particular, findings show that silence pause rate and duration correlate with first language performance while speech rate and filled pause rate do not. Interestingly, repeats also differ from first language production. Results show that the corpus may be a useful tool for researchers who wish to investigate the correspondence between first and second language speech, particularly with respect to the use of hesitation phenomena.

Background

Hesitation phenomena [9,11] may be categorized as follows. One problem with many of these studies is they do not account for first language speech behavior [cf., (6)]. A learner who pauses frequently in L1 might do so in L2, regardless of proficiency. The goal of this study is to provide a resource for the study of L2 hesitation patterns with respect to L1 speech.

Corpus Design

Recording

Studies of hesitation phenomena in second language (L2) speech production have observed several correlations with L2 proficiency [5, 13,14,16,17].

Results

The transcription process is ongoing. Results presented here represent n=18 native Japanese participants. Results show that speech rate and filled pause rate in L2 speech are not correlated with L1 speech patterns (contra [3,7]) while silent pause rate and duration are correlated consistently with [7,9]. Furthermore, results show cross-linguistic differences in the use of repairs, repeats [rare in Japanese [8]], and overall use of hesitation phenomena, but not in ways that are indicative of L2 proficiency development.

Corpus Distribution

The recordings and transcriptions are freely available via an online archive (http://filledpause.com/chp/chcp) under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. Teachers and researchers may make use of the corpus for research and educational purposes.

Conclusions

The preliminary results from the Crosslinguistic Corpus of Hesitation Phenomena show that it can be a useful tool to investigate the relationship between first language speech behavior and second language speech performance, by evaluating the latter with respect to the former on an individual basis. Current results suggest that as second language learners develop higher proficiency in the second language, they speak faster and use fewer filled pauses in a manner that is not necessarily related to their first language rate of speech. On the other hand, their use of other hesitation phenomena—including silent pauses—as their proficiency develops may be more closely related to their first language speech patterns. Future work includes annotation of discourse and clause structure, part-of-speech mark-up, and syllable detection for the purpose of examining leninghens.

Acknowledgments

This research has been funded by a Waseda University Grant for Special Research Projects (#21811-152) and a Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT) Grant-in-Aid (#24520661). I am also grateful to Hiroaki Suzuki, Junichi Inagaki, Masayuki Motoori, Yukikatsu Fukuda, and Tatsuhiro Nomaguchi for their work as research assistants.

References

[12] Kawai Nasal MKC22-26 Dr-40
[16] Kawai Nasal MKC22-26 Dr-40