



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

Ralph L. Rose¹

¹ Waseda University Faculty of Science and Engineering, Tokyo Japan

rose@waseda.jp

Abstract

There is a growing consensus that there is a need to evaluate second language 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 speech behavior. In particular, findings show that silent pause rate and duration as well as other hesitation phenomena correlate with first language performance while speech rate does 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.

Index Terms: hesitation phenomena, second language speech, corpus

1. Introduction

A close examination of everyday speech by native speakers reveals a high frequency of phenomena we might call speech hesitations—long silent pauses, non-verbal vocalizations like *uh* and *um* in English, as well as repairs and repetitions. Much of this goes unnoticed by interlocutors [15] as speakers use them in conventional and unmarked ways that are consistent with native speaker hesitation patterns and norms.

Learning to speak a second language involves developing a sufficient proficiency in producing target language utterances within time constraints pertinent to the communicative situation. In the early stages, learners will often fail to meet these constraints and will take advantage of various strategies to hesitate while preparing their next utterance. The patterns of their hesitation use at this stage might not be the same as the target language norms and therefore can be quite marked, thus indicating low fluency in the language. However, as their proficiency progresses, their hesitation patterns may become more like those of target language norms, hence more unmarked.

Implicit in many second language proficiency hierarchies (such as the ACTFL Proficiency Guidelines [1]) are distinct scales of proficiency development along various trajectories such as vocabulary use, syntactic structure, and pronunciation. A learner's placement along these trajectories can be used to estimate their proficiency. Thus, knowing the typical trajectory

of learners' use of hesitations could also be quite useful in evaluating what stage learners are in in their second language proficiency development. The present research project aims to construct a corpus of learner speech so that a typical developmental trajectory can be determined for English as a second language learners.

However, this effort is made somewhat difficult by the fact that linguistic development can be highly variable across individuals. When it comes to hesitations, in particular, native speakers have highly variable hesitation patterns [8] which could also show up in their second language speech. There may also be non-linguistic aspects of production and planning which influence both their first and second language speech patterns but which have nothing to do with their second language development. Thus, the current project seeks to account for these difficulties by making it possible to interpret learners' second language speech production with respect to their first language speech production patterns and find the measurable aspects of their second language speech which are independent of their first language speech.

2. Background

The various patterns of hesitating in speech—often referred to collectively as hesitation phenomena—have been studied for several decades now. This section gives a brief overview of this research and how they have been studied in the context of second language development.

2.1. Hesitation phenomena

Hesitation phenomena [11], [16] include the following types.

- Silent pauses – long silent pauses, not including the short pauses associated with breathing, articulation, or junctures
- Filled pauses – non-verbal vocalized pauses (*uh/um* in English, *ano/e-to* in Japanese, and *este* in Spanish)
- Repairs – a sequence of speech which is intended to be understood as a replacement of an immediately preceding sequence of speech (*look at the blue the red one over there*)
- Repeats – immediate repetition of a sequence of one or more words (*LL think that's a good idea*)
- False starts – a sequence of speech which begins an utterance but which is then abandoned (*do you I disagree with that*)
- Lengthenings – the prolongation of one or more segments of a word (*I'll take the blue a-nd the- red ones*)

Speakers may use various other strategies to hesitate when speaking including such conventional expressions as *Well...*, *Let me see...*, and *That's a good question*. However, these are generally not included in the study of hesitation phenomena.

Researchers have observed that speakers tend to hesitate more and longer at major discourse boundaries than at minor discourse boundaries [12], [22]. Furthermore, some have observed differences in the use of filled pause sub-types: Closed syllable filled pauses (*um*) are more likely to be followed by longer silent pauses than open syllable filled pauses (*uh*) [4], [20].

In Levelt's well-known model of speech production and monitoring [13], [14], all hesitation phenomena are considered as overt evidence of production repairs accomplished either overtly (e.g., repairs and false starts) or covertly (e.g., silent/filled pauses and lengthenings).

2.2. Use of hesitation phenomena in second language speech production

During the last decade, more and more researchers have looked at the use of hesitation phenomena by speakers in their second language speech production. Evidence shows that higher proficiency speakers use fewer and shorter silent pauses [7], [18], [23], [24] and in some studies, higher proficiency speakers use fewer filled pauses [19].

However, one limitation of many of these studies is that they have not taken first language speech characteristics into account. For example, a speaker who frequently pauses in their second language speech could be merely exhibiting their individual speech characteristics rather than their second language proficiency. Some recent studies are consistent with this hypothesis, showing that some aspects of second language speech behavior are related to first language speech behavior [5], [9]. In particular, silent pause rate as well as speech rate correlated between first and second language speech.

In order to support further investigation of how first language speech behavior relates to second language speech performance, there is a need for crosslinguistic data sets in which parallel data in L1 and L2 is gathered from each participant. The remainder of this paper introduces and describes in detail an ongoing research project to compile a corpus of such speech data with annotated transcriptions for investigative purposes and for public distribution. This corpus is called the Crosslinguistic Corpus of Hesitation Phenomena (hereafter, CCHP).

3. Design of the Crosslinguistic Corpus of Hesitation Phenomena

The CCHP is part of a three-year project to describe a developmental trajectory for the use of hesitation phenomena in second language proficiency development, and to test whether movement along this trajectory can be facilitated through various pedagogical techniques. This paper deals only with the construction of the CCHP. Details of other aspects of the research project will be described elsewhere.

3.1. Data collection procedure

The raw data for the CCHP are a collection of recordings made with university students who were recruited through advertisement in university bulletin boards. Participants were

recorded individually. After signing a consent form which informed them of the public distribution of the corpus, each participant was asked to make three recordings of about 3-4 minutes each in each of their first and second languages (i.e., Japanese and English, respectively). The elicitation tasks for three recordings were as follows (in the order performed).

- Reading aloud: Participants were given a copy of "The Farm Script" [6] and were asked to read it aloud. They were given no advance preparation time. For the English recording they received the original English version of the script. For the Japanese recording, they received a Japanese translation of the script.
- Picture description: Participants were shown black-and-white pictures or cartoon strips (from [2]) one by one and asked to describe each in turn. This was repeated as often as necessary to fill a 3-4 minute time frame. They were told they could take a few seconds to study each picture or cartoon strip, but were asked to begin speaking as soon as possible.
- Topic narrative: Participants were given a topic to talk about freely (e.g., the sport of basketball). They were asked to imagine that they were speaking to someone during this task. If necessary, a second topic (e.g., table tennis) was given to fill a 3-4 minute time frame.

The participants were recorded in a sound-attenuated room using an AKG C300 microphone channeled through an ART Dual Pre microphone pre-amp to a Toshiba Dynabook R731 in mono 16-bit 48kHz quality. The files were processed using the normalize and noise reduction functions in Audacity (ver. 2.0.1; <http://audacity.sourceforge.net/>).

3.2. Transcription procedure

Each recording was transcribed for spoken word and partial word tokens and annotations were made for filled pauses (most commonly, *uh/um* in English, *e-to/ano-* in Japanese), false starts, the structure of repair sequences (i.e., reparandum, editing terms, and repairs; cf., [13], [21]), and a few other minor audible phenomena (e.g., coughs, throat-clearing, non-verbal interjections like "ah!"). Each recording was processed by two transcribers independently (the inter-transcriber agreement is 91.8%, an acceptable rate, cf., [17]) and differences were resolved by a third checker. Pause and word interval durations were detected using the default pause/speech detection script in Praat [3] and then manually checked. Transcripts are stored in XML format and audio recordings in wav format.

Following is a short extract comprising one utterance from one transcription. A repair sequence is indicated by <RP>, the reparandum by <O>, the repair by <E>, and editing terms as nodes between <O> and <E>.

```
<UTTERANCE>
<T>in</T>
<T>America</T>
<T FILLED-PAUSE="yes">uh</T>
<T>there's</T>
<T>a</T>
<T FILLED-PAUSE="yes">uh</T>
<T>very</T>
<T>famous</T>
<T FILLED-PAUSE="yes">uh</T>
<T>and</T>
```

```

<T>loved</T>
<T FILLED-PAUSE="yes">uh</T>
<T>basketball</T>
<RP>
<O>
  <T>cl#</T>
</O>
<T FILLED-PAUSE="yes">uh</T>
<E>
  <T>association</T>
</E>
</RP>
<T>which</T>
<T>is</T>
<T>called</T>
<T>NBA</T>
<T>National</T>
<T>Basketball</T>
<T>Association</T>
<T>I</T>
<T>think</T>
</UTTERANCE>

```

3.3. Demographic information

Some demographic information about each participant was collected to assist the interpretation of the participants' second language speech characteristics. This included age, gender, experience living abroad, self-estimate of foreign language ability, and results of English language proficiency tests.

3.4. Public availability

The recordings and transcripts (but not the demographic information, for privacy reasons) are freely available via an online archive (<http://filledpause.com/chp/cchp>) 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.

4. Results

Recordings were made during a 10-month period from June 2012 to March 2013 with 35 participants. Some basic statistics on the number of tokens (words plus filled pauses) and duration are shown in Tables 1 and 2 while the counts of various hesitation phenomena are shown in Table 3. [Note: The transcription process is still ongoing, so this data represents only 15 of the 35 participants.]

Table 1. *Token (words, filled pauses) count*

	Reading aloud	Picture description	Topic narrative	Total
Japanese	4,246	4,375	5,086	13,707
English	4,897	2,960	2,637	10,494

Table 2. *Overall duration (average duration per participant is shown in parentheses)*

	Reading aloud	Picture description	Topic narrative	Total
Japanese	31.1 min (124.5 sec)	56.6 min (226.4 sec)	56.3 min (225.2 sec)	144.0 min
English	39.4 min (157.6 sec)	61.9 min (247.6 sec)	58.2 min (233.0 sec)	160.0 min

Table 3. *Overall count of various hesitation phenomena*

	Japanese	English
Silent pauses	3,106	3,841
Filled pauses, total	742	535
Open type (<i>uh</i>)	572	324
Closed type (<i>um</i>)	170	211
Repair sequences	231	348
Repeats	28	149

Analysis of the corpus reveals some interesting results. Although speech rate is not a type of hesitation phenomena, it is a related temporal variable and is useful to examine in the same context. Figure 1 shows the relationship between speech rate and second language proficiency (estimated from demographic information and with a range covering the novice to superior levels in the ACTFL Proficiency Guidelines [1]).

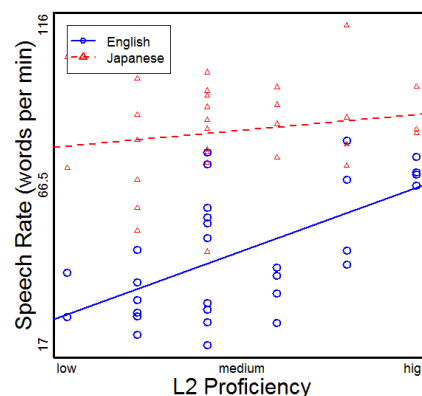


Figure 1. *Speech rate and second language proficiency*

Both main factors are significant as is the interaction between them [$F(1,13)=4.7, p<0.05$]. These results are consistent with earlier studies showing that speech rate corresponds well with increased second language proficiency yet further shows that speech rate may be a reliable predictor independently of first language speech performance. However, they are not consistent with recent findings that speech rate is highly correlated between first and language speech production [5], [9]. Further work is necessary to explore this inconsistency.

The results for silent pause rate and for silent pause duration are shown in Figure 2 and Figure 3, respectively.

For both silent pause rate and duration, there are no significant main effects or interaction. These results are consistent with recent studies showing a correlation between silent pause use in first and second language speech performance [5], [9].

Another interesting result is observed with repeats. Figure 4 shows the rate of repeats with respect to second language proficiency.

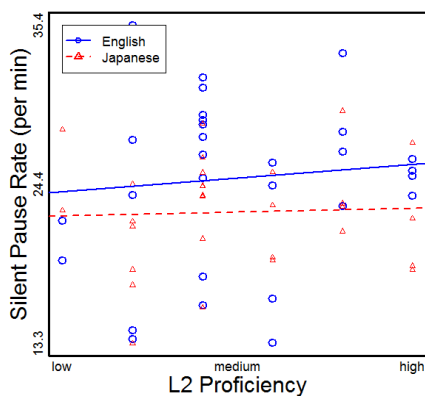


Figure 2. Silent pause rate and second language proficiency

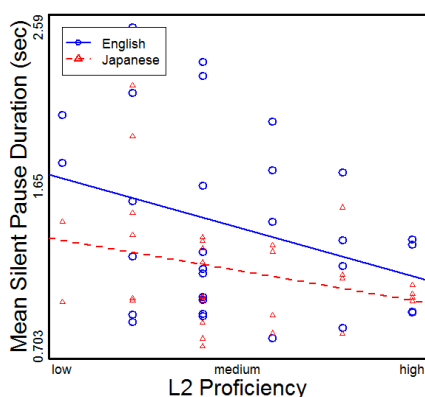


Figure 3. Mean silent pause duration and second language proficiency

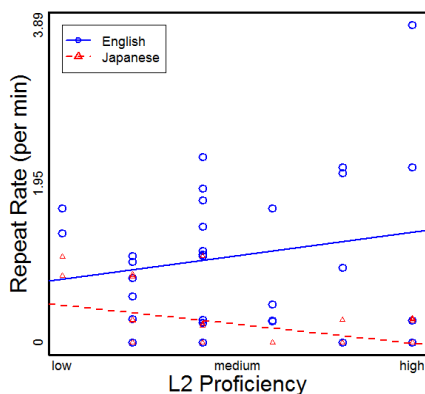


Figure 4. Repeat rate and second language proficiency

There is a main effect of language [$F(1,13)=17.3$, $p<0.005$], but no other main effect or interaction. The occurrence of repeats in Japanese has been observed to be infrequent [10]. The results in Figure 4 are consistent with this observation. The data further suggests that as speakers become more proficient in English, their use of repeats increases, though this trend is not significant.

5. Discussion

The preliminary results from the CCHP 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, 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.

These results have implications for the evaluation of second language proficiency either by human evaluators or automated agents. Preliminary results suggest that only speech rate would be a reliable predictor, though deeper analysis might be useful to see if some particular combination of factors could yield more promising results.

6. Conclusions

This paper has described the CCHP in detail and presented some preliminary results which are consistent with previous findings in showing that L2 speech performance depends on L1 speech behavior. Future work includes annotation of discourse and clause structure, part-of-speech mark-up, and syllable detection for the purpose of examining lengthenings. The freely-available CCHP should be of interest and use to those who wish to study hesitation phenomena or who are doing other research related to temporal phenomena in speech production (e.g., fluency, speaker variation). The crosslinguistic and parallel structure of the corpus should be of particular interest to those who wish to look at L2 speech performance data in comparison to L1 speech behavior.

7. Acknowledgements

This research has been funded by a Waseda University Grant for Special Research Projects (#2011B-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, and Yukikatsu Fukuda for their work as research assistants.

8. References

- [1] ACTFL Proficiency Guidelines 2012, American Council on the Teaching of Foreign Languages, 2012.
- [2] Alexander, L. G., First Things First, Longman, 1967.
- [3] Boersma, P., "Praat, a system for doing phonetics by computer", *Glott International*, 5(9/10):341-345, 2001.
- [4] Clark, H. and Fox Tree, J. E., "Using uh and um in spontaneous speaking", *Cognition*, 84(1):73-111, 2002.
- [5] Cox, T. and Baker-Smemoe, W., "The relationship between L1 fluency and L2 fluency across different proficiency levels and L1s", presentation at Workshop Fluent Speech (Utrecht University, The Netherlands), 2012.
- [6] Crystal, T. H., & House, A. S., "Segmental durations in connected speech signals: Preliminary results", *Journal of Acoustical Society of America*, 72(3):705-716, 1982.
- [7] Cucchiari, C., van Doremalen, J. and Strik, H., "Fluency in non-native read and spontaneous speech", *Proceedings of Disfluency in Spontaneous Speech (DiSS) and Linguistic*

Patterns in Spontaneous Speech (LPSS) Joint Workshop, pages 15-18, 2010.

- [8] de Leeuw, E. "Hesitation markers in English, German, and Dutch", *Journal of Germanic Linguistics*, 19(2):85-114, 2007.
- [9] Derwing, T. M., Munro, M. J., Thomson, R. I. and Rossiter, M. J., "The relationship between L1 fluency and L2 fluency development", *Studies in Second Language Acquisition*, 31(4):533-557, 2009.
- [10] Fox, B., Hayashi, M., and Jaspersen, R., "Resources and repair: A cross-linguistic study of syntax and repair". In Ochs, E., Schegloff, E., and Thompson, S. (eds). *Interaction and Grammar*. Cambridge University Press, *Studies in Interactional Sociolinguistics*, volume 13: pages 185-237, 1996.
- [11] Goldman-Eisler, F., "A Comparative Study of Two Hesitation Phenomena", *Language and Speech*, 4(1):18-26, 1961.
- [12] Goldman-Eisler, F., "Pauses, Clauses, Sentences", *Language and Speech*, 15(2):103-113, 1972.
- [13] Levelt, W. J. M., "Monitoring and self-repair in speech", *Cognition*, 14(1):41-104, 1983.
- [14] Levelt, W. J. M., *Speaking: from intention to articulation*, MIT Press, ACL-MIT Press series in natural-language processing, 1989.
- [15] Lickley, R. J., "Missing Disfluencies", *Proceedings of International Congress of Phonetic Science*, 192-195, 1995.
- [16] Maclay, H. and Osgood, C., "Hesitation Phenomena in Spontaneous English Speech", *Word*, 15:19-44, 1959.
- [17] Patterson, B. R., Neupauer, N. C., Burant, P. A., Koehn, S. C., and Reed, A. T., "A preliminary examination of conversation analytic techniques: Rates of inter-transcriber reliability", *Western Journal of Communication*, 60(1):76-91, 1996.
- [18] Riazantseva, A., "Second language proficiency and pausing: A study of Russian speakers of English", *Studies in Second Language Acquisition*, 23(4):497-526, 2001.
- [19] Rieger, C. L., "Disfluencies and hesitation strategies in oral L2 tests", *Proceedings of DiSS '03, Disfluency in Spontaneous Speech Workshop*, 2003.
- [20] Rose, R. L., "The Information Value of Some Hesitation Phenomena: Filled Pauses, Lengthenings, and Entropy Reduction", poster presented at the Japan Society for Linguistic Sciences (JLS) Annual Meeting (Saitama, Japan), 2009.
- [21] Shriberg, E., *Preliminaries to a theory of speech disfluencies*, doctoral dissertation, University of California, Berkeley, 1994.
- [22] Swerts, M., "Filled pauses as markers of discourse structure", *Journal of Pragmatics*, 30(4):485-496, 1998.
- [23] Tavakoli, P., "Pausing patterns: differences between L2 learners and native speakers", *ELT Journal*, 65(1):71-79, 2011.
- [24] Trofimovich, P. and Baker, W., "Learning prosody and fluency characteristics of second language speech: The effect of experience on child learners' acquisition of five suprasegmentals", *Applied Psycholinguistics*, 28(2):251-276, 2007.