On the Relationship between First and Second Language Spontaneous Speech: The Role of Hesitation Phenomena

Ralph L. Rose
<rose@waseda.jp>

Center for English Language Education (CELESE)
Waseda University Faculty of Science and Engineering

Acknowledgments
Hiroaki Suzuki, Junichi Inagaki, Masayuki Motoori,
Yukikatsu Fukuda, & Tatsuhiro Nomaguchi

AILA 2014, Brisbane
August 10-16

This research is partially supported 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)
Introduction

Well, i- i- i- in my own life I'd- I'd break it up in stages, when uh I had a difficult youth. Uh my father wasn't in the house, uh I've written about this, uh there- uh uh you know there were times where uh I've experimented with drugs, and I drank, uh yeah in my teenage years, and what I trace this to is uh a certain selfishness on my part, I- I was so obsessed with me, and you know the reasons that I might be dissatisfied, that I- I- w- I couldn't focus on other people. And uh y- you know I think the process for me of growing up was to recognize that it's not about me, it's about ...

it's about- absolutely, so- so- but- but look, you know, th- the uh wh- when I uh wh- when I find myself um taking the wrong step, I think a lot of the times it's because I'm trying to protect myself, instead of trying to do God's work. And- and- an- and so that I think is- is my own failure

Barack Obama (August, 2008); Saddleback Presidential Forum
Overview

- Hesitation phenomena in speech
  - Overview
  - In L2 speech
- Crosslinguistic Corpus of Hesitation Phenomena
  - Description
  - Results
- Developmental Trajectory
- Accessing the CCHP
Hesitation phenomena

Silent pauses longer than 0.3-1.0sec

Self-corrections (repairs)
Sequence that repairs a preceding sequence
*Look at the blue the red one over there.*

Filled pauses *uh/um* (English)
*e-to/ano-* (Japanese)

Lengthenings
Prolongation of one or more syllables
*I'll take the blue *and* the red ones.*

Repeats/Repeats
Repetition of a sequence of words
*I I I I think that's a good idea.*

False starts
Beginning of utterance that is abandoned
*Do you I disagree with that.*

Speech rate by word, by syllable, with/without pauses
*(Cucchiarini et al 2010)*

Hesitation phenomena in L2 production

Hesitation phenomena in L2 production

- As a whole, work has been quite comprehensive.
- However, individual works are limited in that many do not take individual variation into account (cf., de Leeuw 2007).
- Gradually, more studies are including L1 observations.
  - Derwing et al (2009) and Cox and Baker-Smemoe (2012) observed that both speech rate and pause rate in L1 and L2 production are correlated.
- The current research is designed to contribute to greater understanding of the relationship between L1 hesitation patterns and L2 hesitation patterns.
Research questions

• What is the relationship between hesitation patterns in L1 and L2 speech?
• What is the developmental trajectory of the use of hesitation phenomena in L2?
Crosslinguistic Corpus of Hesitation Phenomena (CCHP)

- Participants: L2 learners of varying proficiency levels
- Elicitation tasks
  - Spontaneous speech: picture description, topic narrative
  - Reading aloud
  - Performed in both L1 and L2
Crosslinguistic Corpus of Hesitation Phenomena (CCHP)

- Demographic information: age, gender, L2 proficiency (standardized test scores, experience abroad, self-assessment)
- Annotation
  - Transcripts, HP, word & pause intervals
  - Two annotators, one checker

<UTTERANCE>
<T>in</T> <T>in</T> <T>America</T> <T>America</T> <T FILLED-PAUSE="yes">uh</T> <T FILLED-PAUSE="yes">uh</T> <T>there's</T> <T>there's</T> <T>a</T> <T>a</T> <T FILLED-PAUSE="yes">uh</T> <T FILLED-PAUSE="yes">uh</T> <T>very</T> <T>very</T> <T>famous</T> <T>famous</T> <T FILLED-PAUSE="yes">uh</T> <T FILLED-PAUSE="yes">uh</T> <T>and</T> <T>and</T> <T>loved</T> <T>loved</T> <T FILLED-PAUSE="yes">uh</T> <T FILLED-PAUSE="yes">uh</T> <T>basketball</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>which</T> <T>is</T> <T>is</T> <T>called</T> <T>called</T> <T>NBA</T> <T>NBA</T> <T>National</T> <T>National</T> <T>Basketball</T> <T>Basketball</T> <T>Association</T> <T>Association</T> <T>I</T> <T>I</T> <T>think</T> <T>think</T> </UTTERANCE>
CCHP Results: Basic Statistics

- Participants: 36 Japanese L1 / English L2 speakers
- Full corpus
  - 62,632 words
  - 11 hrs, 31 min
- Spontaneous speech
  - 40,296 words
  - 8 hrs, 43 min
- Read speech
  - 22,336 words
  - 2 hr, 48 min
- Transcriber agreement
  - 91.5%

- 15,837 silent pauses
- 3,516 filled pauses
- 1,689 self-corrections
- 518 repeats
CCHP Results: Hesitation Index

Hesitation Index

\[
\text{Hesitation Index} = 1 - \frac{\text{Number of essence words}}{\text{Number of spoken words}}
\]

Where “essence” is what the speaker intended to say.
CCHP Results: Speech Rate

Consistent with Derwing et al (2009) and Cox and Baker-Smemoe (2012)
CCHP Results: Silent Pauses

Silent Pause Rate

Silent Pause Duration

CCHP Results: Filled Pauses

Filled Pause Rate

Interaction

L2 Proficiency

FP Rate (per 100 words)
CCHP Results: Self-corrections

Self-correction Rate

Repair Rate (per 100 words)

L2 Proficiency

English
Japanese

low medium high
CCHP Results: Repeats

Repeats are uncommon in Japanese (Fox et al 1996)
Developmental Trajectory

To-do:
- Filled pause duration
- Lengthenings
- Repair type distribution
- Structural distribution
- Syllable counts
Summary

- Recent studies of L2 speech performance are taking L1 speech performance more and more into account.
- The Crosslinguistic Corpus of Hesitation Phenomena allows us to account for L1 factors in the study of L2 speech patterns.
- Results show that silent and filled pause features indicate learners' L2 proficiency.
- Results suggest that other L2 hesitation phenomena correlate with those of L1.
CCHP Public Corpus

- Assembling a public version of the Crosslinguistic Corpus of Hesitation Phenomena is ongoing.
- When complete, audio files and annotated transcripts will be available for free download.
- Some files are already available for download: http://www.filledpause.com/chp/cchp
References


