

# Final Project Presentation

## Analysis: Non-native Speech & SDS



*Marie-Renée Arend*

5 June 2013

# The questions I wanted to answer

In spoken dialogue systems:

What are the qualitative differences in errors between native and non-native speakers?

What kinds of solutions are effective in preventing or correcting errors caused by non-native speech?

Credit due to: Michael Tjalve & Jason Williams at MSR for helping me find a direction for my analysis and guiding me toward helpful resources

# Why do those questions matter?

## From the user's perspective:

- Main user population might be non-native speakers
  - Ex: tourist info, social svc info, legal info
- Incorporation of non-native data into LM has been shown to also increase recognition for native speakers too – everyone wins!
- Systems for users in “global” cities where population has a large variety of linguistic backgrounds
  - Ex: problems with creating a recognizer for inhabitants of Hong Kong

## From the system designer's perspective:

- Depending on the product:
  - Could want to make the system usable for as many speaker populations as possible
  - Could focus on excellent recognition of non-native speech from speakers w/ the same native language
- If creating an adaptive system, don't want to “scare off” non-native users from using it; need their data to improve model – but why would someone use a SDS that doesn't understand them?
- Improve error detection and handling with minimal effort from the user to get the task done more efficiently

# Differences between native & non-native speech characteristics that can lead to errors

## ☞ Native speech:

- Disfluencies
- Lexical variants (mainly OOV words)
- Prosodic variants
  - Speaking rhythm, stress (particularly hyperarticulation)

## ☞ Non-native speech:

- Disfluencies
  - I.e., abandoned words, stutters, restarts, filler words, etc
- Adding in words from their own language
- Pronunciation variants
- Lexical variants/creativity
- Syntactic variants
- Excessive pausing
- Prosodic variants

# Solutions

## ∞ Linguistic mismatch (variations/errors in syntax or vocab)

- Dialogue context cues/information
- Adaptive lexical entrainment
  - Ex: confirmation module which stresses the word that the user missed when asking the user what he/she meant to say
- Augmentation of LM with non-native data
- Flexibility of grammar/vocabulary
- Parser robustness

## ∞ Acoustic mismatch (pronunciation variants, accents, etc)

- Speeding up the speech rate at the feature level to a native speaker's typical pace
- Modifying the pronunciation dictionary to allow for some variants
  - Can be done either for specific TLs or generalized for multiple TLs
- Rule-based modification in a post-processing stage – can be applied to linguistic mismatch as well
  - Handwritten rules based on linguistic knowledge
  - Data-driven extraction of patterns from corpora

# Discussion

Why so many different ways to go about solving this problem?

- ∞ Availability of training data for the SDS
  - Often, data can be drawn from only a small population of non-native speakers
- ∞ Capability of the system to adapt to individual users
  - Can the system “learn” from its non-native users to better understand them?
- ∞ General system design
  - Ex: how flexible is the grammar in allowing for syntactic variants?
- ∞ The purpose of the system (i.e., the task it is meant to assist users with)
  - Ex: computer-assisted language learning systems vs. Let’s Go! bus system
- ∞ Domain
- ∞ Creating a system for non-native speakers with the same native language vs. a system that works for non-native speakers with diverse linguistic backgrounds

# Resources

- ☞ Beňuš et al. 2011: “Adapting Slovak ASR for native Germans speaking Slovak”
- ☞ Fung et al. 1997: “Dealing with Multilinguality in a Spoken Language Query Translator”
- ☞ Lahiri & Mihalcea, 2013: “Using N-gram and Word Network Features for Native Language Identification”
- ☞ Litman et al. 2006: “Characterizing and Predicting Corrections in Spoken Dialogue Systems”
- ☞ Raux, 2004: “Automated lexical adaptation and speaker clustering based on pronunciation habits for non-native speech recognition”
- ☞ Raux & Eskenazi, 2004: “Non-native users in the Let’s Go!! spoken dialogue system: dealing with linguistic mismatch”
- ☞ Raux et al. 2003: “Let’s Go: Improving spoken dialogue systems for the elderly and non-natives”
- ☞ Schaden, Stefan. “Rule-based lexical modelling of foreign-accented pronunciation variants”
- ☞ Tomokiyo, 2000: “Linguistic properties of non-native speech”
- ☞ Xu & Seneff 2012: "Improving Nonnative Speech Understanding Using Context and N-Best Meaning Fusion"

# Questions?



*Thanks for listening!*