The Canadian multilingual speech database
Yasaman Rafat, Western University & Rallis Karamichalis, Western University

The main aim of this project is to create a Canadian multilingual and multigenerational speech database. There has been a dramatic interest in research on second/foreign language speech learning (e.g., Flege, 1995; Colantoni & Steele, 2008, Rafat, 2016), second dialect acquisition (e.g., MacLeod, 2012; Ruiz-Pena, Sevilla & Rafat, 2015), first language phonetic and phonological attrition and change (e.g., Optiz, 2013; Rafat & Mohaghegh, 2015; Schmid, 2014; Ulbrich & Ordin, 2014), heritage languages (e.g.,Chang, Haynes & Rhodes, 2011; Rafat & Mohaghegh, 2015), as well as foreign accent (e.g., Boula de Mareüil, Vieru-Dimulescu, 2006; Wieling, Bloem, Mignella, Timmermeister, & Nerbonne, 2014). Although there is an American Speech Accent Archive (Weinberger, 2014) (http://accent.gmu.edu/) and an Australian National Database of Spoken Language (Millar, Vonwiller, Harrington & Dermody, 1994) (http://andosl.rsise.anu.edu.au/andosl/), we do not currently have a Canadian database that represents both the English and the native/heritage languages of immigrants in Canada and is publicly available to research community and those interested in languages. Canada is a multicultural immigrant receiving country, where both second language learning and preserving heritage languages are highly relevant, and valued. The creation of such a database is important for a number of reasons. First, it allows scholars to work on language change in a contact situation and understand how different varieties of English spoken evolve over the years and how native/heritage languages change in a contact situation. Second, this project will documents the languages of the immigrant population in Canada, preserving our unique heritage cultures that make up the Canadian multicultural mosaic and identity. Third, the database will be used as a pedagogical tool in both schools and universities for teaching ESL, heritage languages and foreign accent reduction.

We are in the first phase of the project, where speech samples are being collected and documented of the native/heritage languages and English from first, second and third generation bilingual and multilingual immigrants living in Ontario. We will include different groups such as, Spanish-French-English, Italian-French-English, Mandarin-Cantonese-English, Farsi-Azeri/Turkish-English multilinguals and Hindi-Punjabi-bilinguals. The data will include a) an informal sociolinguistic interview in each of the languages they speak, (b) a sentence reading task in English and their other (heritage/native) language(s), and (c) a sociolinguistic background questionnaire. Detailed anonymous descriptors of the speakers will be available as part of the data corpus description. The database is built using PHP a widely-used open source scripting language, and MySQL an open source relational database management system. The audio recordings produced will be licensed under a Creative Commons License, and more specifically under Attribution-NonCommercial-ShareAlike 2.0 Generic (CC BY-NC-SA 2.0) (https://creativecommons.org/licenses/by-nc-sa/2.0/). The database will also be connected to social media. Researchers will be able to download data (e.g., sound files) for free and without having to contact the owners. We hope that the database opens up possibilities for collaboration both within and outside the university, where it is being

built. This is work in progress and we are planning on expanding our data collection to other regions of Canada.

References

Boula de Mareüil, P., Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of foreign accent. *Phonetica*, 63(4), 247-67

Colantoni, L., & Steele, J. (2008). Integrating articulatory constraints into models of second language phonological acquisition. *Applied Psycholinguistics*, 29(03), 489-534.

Chang, C. B., Yao, Y., Haynes, E. F., & Rhodes, R. (2011). Production of phonetic and phonological contrast by heritage speakers of Mandarin. *The Journal of the Acoustical Society of America*, 129(6), 3964-3980.

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Effects of age of second-language learning on the production of English consonants. Speech Communication, 16(1), 1-26.

Millar, J. B., Vonwiller, J. P., Harrington, J. M., & Dermody, P. J. (1994, April). The Australian national database of spoken language. In *Acoustics, Speech, and Signal Processing*, 1994. ICASSP-94., 1994 IEEE International Conference on (Vol. 1, pp. I-97). IEEE.

Opitz, C. (2013). A dynamic perspective on late bilinguals' linguistic development in an L2 environment. *International Journal of Bilingualism* 17:6

Ruiz-Peña, E., Sevilla, D., & Rafat, Y. (2015). Second dialect imitation: The production of Ecuadorian Spanish assibilated rhotics by Andalusian speakers of Spanish. In E. Babatsouli & D. Ingram (eds.), *Proceedings of the International Symposium on Monolingual and Bilingual Speech* 2015 (pp. 288-300). ISBN: 978-618-82351-0-6. URL: http://ismbs.eu/publications

Rafat, Y. & Mohaghegh, M. Geminate consonant production across three generations of Farsi-English-speaking Iranian-Canadian bilinguals. *The Canadian Linguistics Association Annual Conference*, University of Ottawa, Ottawa, ON. (May 2015)

Rafat, Y. (2016). Orthography-induced transfer in the production of English-speaking learners of Spanish. *The Language Learning Journal*, 44(2), 197-213.

Schmid, M.S. (2014) The debate on maturational constraints in bilingual development: A perspective from first language attrition. *Language Acquisition* 21, 386-410.

Ulbrich, C. & M. Ordin. (2014). Can L2-English influence L1-German? The case of post-vocalic /r/. *Journal of Phonetics* 45, 26-42.

Weinberger, S. H. (2014). Speech Accent Archive. George Mason University. Online:< http://accent. gmu. edu.

Wieling, M., Bloem, J., Mignella, K., Timmermeister, M., & Nerbonne, J. (2014). Measuring foreign accent strength in English. *Language Dynamics and Change*, 4(2), 253-269.