Methodological Approaches to Multilingual Annotated Corpus Design
Abbie Hantgan, SOAS

This paper illustrates methodological approaches to the design of an annotated corpus using a case study of phonetic convergences and divergences by multilingual speakers in South-western Senegal. Each of three villages which is the focus of the study is associated with a patrimonial language, and many more languages contribute to the area's language landscape and thus to individuals' repertoires. Included in the corpus which is the focus of this presentation are data from each of the three villages which were gathered using a wide variety of methods ranging from structured lexical elicitation in controlled, monolingual contexts to naturally occurring, multilingual conversations collected as part of an on-going project's social network study. Participants in the social network study are recorded throughout their normal daily activities through the use of a portable digital recorder. Their conversations are segmented by intonational unit, transcribed and translated by trained members of the community and annotated by project researchers. Because of detailed metadata included within each of the corpus' speech sessions, we are able to discern the context in which a speech act took place and between whom, and therefore underpin analyses of variation with detailed information on speakers' linguistic repertoires, trajectories, social networks as well as on the language context (see abstract "Multilingual corpora for differentiated approaches to language contact" for details on language contexts).

An area which has emerged during the researcher's data collection and through preliminary searches of the existing corpus as a whole in which speakers of the area have the potential to converge or diverge is in the phonetic production of certain consonants, namely velars. Speakers' potential to converge or diverge their pronunciation of velar consonants relies on differences among the prototypical phonetic representation of velars in each of the villages' patrimonial languages.

Evidenced by minimal pairs from Jóola Kujireray, the language identified with the village Brin, and Jóola Bandial, the language spoken six kilometres away in the village of Essyl, the sounds [k] and [g] are mutually exclusive. In Baïnounk Gubëeher, the language associated with Brin's neighbouring village Djibonker, the sounds are phonemically contrastive. Based on this evidence alone, we would expect speakers of Jóola Kujireray to use [k] in situations where speakers of Jóola Bandial pronounce [g]. Indeed, in emblematic contexts such as greetings, this is exactly occurs:

Jóola Kujireray	Jóola Bandial	Gloss
kasuumaj	gəəssumaj	peace

For the moment concentrating only on the word-initial segment, we see that speakers of the adjacent villages pronounce a form that is structurally similar in the two languages differently. Further, speakers of Baïnounk Gubëeher, when speaking in one of the Jóola languages, are flexible in their speech and are able to align themselves with either of the two villages' pronunciation. However, what is somewhat unexpected given the underlying representations of the consonants in the two mutually unintelligible languages is Jóola Kujireray and Jóola Bandial speakers' capacity to converge or diverge their pronunciation of velars, even in the most emblematic of contexts such as greetings.

While many studies have examined the effect on relative VOT of bilinguals who have been living in a region different than that in which they grew up, (Flege and Eefting 1987, Sancier and Fowler 1997, Fowler et al. 2008, Ayala 2011), few (Beyer 2015) have examined the situations which lead multilingual speakers' to either converge or diverge their pronunciation to that of their interlocutors. Due to the project's corpus design, not only is it feasible to identify areas of phonetic convergence and divergence, it is also possible to gain an understanding of the context surrounding the event and in turn propose ethnographically and sociolinguistically informed analyses of the speakers' linguistic behaviour based on their backgrounds and repertoires. In addition, through identifying for which speakers and in which language contexts divergences and convergences are most and least frequent, we can zoom in on the most dynamic and most conservative areas of language use, thus contributing to the development of predictive models of language change in multilingual settings.

References

Ayala, A. (2011). *Phonetic Convergence: A Case Study of a Puerto Rican Spanish Speaker* (Senior Essay). Yale University.

Beyer, K. (2015). Multilingual speakers in a West-African contact zone: An integrated approach to contact-induced language change. In Stell, Gerald and Yakpo, Kofi (Ed), *Code-switching Between Structural and Sociolinguistic Perspectives* (237 - 258). De Gruyter Mouton

Flege, J., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, *6*(3), 185 - 202.

Fowler, C. A., Sramko, V., Ostry, D. J., Rowland, S. & Halle, P. (2008). Cross language phonetic influences on the speech of French-English bilinguals. *Journal of Phonetics*, 36, 649 - 663.

Sancier, M.L. & Fowler, C.A. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25, 421- 436.