Multilingual corpora for differentiated approaches to language contact

Friederike Lüpke

SOAS, University of London

fl2@soas.ac.uk

The ongoing Crossroads project (http://soascrossroads.org/) investigates rural multilingualism in Southern Senegal (West Africa). There, speakers have a repertoire of between 4 and 10 named languages. We have collected data from a high number of speaker and different types of communicative events (see Abbie Hantgan's abstract for a description of the corpus design and a pilot case study on phonetic variation). These data have been enriched by detailed ethnographic and sociolinguistic metadata on the demographic details, self-reported repertoires and social relationships of research participants. Through participant observation and a preliminary qualitative corpus study, we have developed first hypotheses on different contexts of multilingual language use.

While 'languaging' (García & Wei 2013) is widespread, named languages as ideological constructs offer a connection to the land. The patrimonial language is the (remembered) language of the founding clan of a village, or to be more precise, in the mostly patrilineal and virilocal societies of the area, its male members. Claiming the patrimonial language (which not always entails speaking it), lends autochthone status and conveys land rights. Rather than linguistically assimilating strangers, landlords encourage the strangers they host to keep the linguistic identities of their villages of origin. Societies in the area are often (linguistically) exogynous. Since women move into their husbands' households upon marriage, their versatile repertoires contribute to weaving multilingualism deeply into the fabrics of society. I illustrate the dualism manifest in the opposition between flexible language use and ideologically fixed language with corpus-based examples of attested multilingual patterns, ranging from more monolingual to thickly multilingual modes. Following Green & Abutalebi (2013), we have identified a 'single language context', present when participants share one main language. This context is close to the patrimonial language prototype that is never fully instantiated in discourse. Linguistically, it is characterised by insertional code-switching (Auer 1999) and the most conventionalised contact patterns. The, 'dual language context' (Green & Abutalebi 2013) can contain two or more named languages, with each code used with a different interlocutor. Here, we find *en-bloc* code-switching. Finally, a third 'intense code-switching context (Green & Abutalebi 2013) is located close to the 'fused lect' end of Auer's (1999) code-switching cline and exhibits intense and variable mixing, containing many lexico-grammatical convergences. I present speech of these three contexts, using the distance from the monolingual prototype to calibrate the convergences, and providing pilot quantitative data on the frequencies of the context and of contact phenomena contained in them. Finally, I emphasise the potential of multilingual corpora to obtain a nuanced picture of language contact. Rather than generalising over entire speech communities, corpora allow differentiating types of communicative situations within one community, singling out the most monolingual as well as the most multilingual modes and permitting a contrastive study of contact patterns in them. Such an approach increases the predictive power of contact studies, since it can elucidate which patterns occur in particular situations, and how the relative frequency of monolingual vs. multilingual interaction of particular types contributes the dynamics of language change.