Workshop “Corpus-driven studies of heterogeneous and multilingual corpora”

Where: CNRS campus at Villejuif, France  
When: September 26, 2016

Convenors: Evangelia Adamou and Isabelle Léglise

In the past ten years, there has been a growing number of spoken corpora annotated for language contact including for lesser-described languages (Adamou 2016) and for multilingual settings illustrating all sorts of bilingual talk, codeswitching, heterogeneous language practices or polylanguaging (Léglise & Alby 2013, 2016).

This workshop will focus on the quantitative analysis of free-speech corpora in various settings and what they bring for the description of mixed languages, bilingual talks and mixed or heterogeneous language practices. Indeed, a quantitative approach of spoken corpora has demonstrated the creation of special outcomes of codeswitching, such as “mixed languages” (McConvell & Meakins 2005), and so-called “unevenly mixed languages” (Adamou & Granqvist 2015). Moreover, the quantitative approaches of mixed and heterogeneous corpora of typologically diverse languages can re-evaluate the relevance of the very influential “matrix” and “numerically-dominant language” concepts (Myers-Scotton 1993, Muysken 2000).

More specifically, this workshop aims at bringing together researchers addressing the following questions:

- The degrees of language mixing based on a quantitative analysis of spoken corpora.
- The degrees of heterogeneity within corpora.
- Typologies of bilingual talk.
- Language mixing typologies.

Invited speaker: Felicity Meakins (University of Queensland, Brisbane), “Breadth and depth: Quantifying complexity across populations in situations of language contact”.

Funding: 
This workshop is part of the project “Multifactorial analysis of language changes” http://axe3.labex-efl.org/?q=fr/LC1f (PI Isabelle Léglise) funded by the French National Research Agency (ANR) as part of the “Investissements d’Avenir” program (ANR-10-LABX-0083) Empirical Foundations of Linguistics.

References