



Population effects on languages:

*Modelling population dynamics and language transmission
from the perspective of language learning, contact and change*

Interdisciplinary Workshop — November, 20 2017 — Lyon, France

<https://poplang.sciencesconf.org/>

HOW SOCIAL, ECONOMIC AND DEMOGRAPHIC FORCES SHAPE LINGUISTIC VARIATION ON TWITTER

**Jacobo LEVY ABITBOL¹ Márton KARSAI¹, Jean-Pierre CHEVROT²,
Jean-Philippe MAGUÉ³, and Éric FLEURY¹**

¹ Université de Lyon, ENS de Lyon, Inria, CNRS, UCBL, F69342, France

² Université Grenoble Alpes, Lidilem, France

³ Université de Lyon, ENS de Lyon, ICAR UMR 5191, France

Our usage of language is not solely reliant on cognition but is arguably determined by a myriad of external factors leading to a global variability of linguistic patterns. This issue, which lies at the core of sociolinguistics and is backed by many small-scale studies on face-to-face communication, is addressed here by constructing a dataset combining the largest French Twitter corpus to date with detailed socioeconomic maps obtained from national census in France. We focused on three individual linguistic markers as (a) the deletion of the first morpheme of the French negation (Nonstandard: *Je fume pas* vs. Standard: *Je ne fume pas* - *I do not smoke*); (b) correct usage of plural forms (Nonstandard: *les livre rouge* vs. Standard: *les livres rouges* - *the red books*); and (c) the vocabulary set size computed for each users. We then completed our dataset by identifying the home locations of around 300,000 users by studying GPS-tagged tweets. We matched the home places with spatially localized socioeconomic data collected during the census in France. We show how these key linguistic variables measured in individual Twitter streams depend on factors like socioeconomic status, location, time, and the social network of individuals.