Exploring British accents: Modelling the trap–bath split with functional data analysis
Shahin Tavakoli, University of Geneva

The sound of our speech is influenced by the places we come from. Great Britain contains a wide variety of distinctive accents which are of interest to linguistics. In particular, the 'a' vowel in words like 'class' is pronounced differently in the North and the South. Speech recordings of this vowel can be represented as formant curves or as mel-frequency cepstral coefficient curves. Functional data analysis and generalised additive models offer techniques to model the variation in these curves. Our first aim was to model the difference between typical Northern and Southern vowels /æ/ and /a/, by training two classifiers on the North-South Class Vowels dataset collected for this paper. Our second aim is to visualise geographical variation of accents in Great Britain. For this we use speech recordings from a second dataset, the British National Corpus (BNC) audio edition. The trained models are used to predict the accent of speakers in the BNC, and then we model the geographical patterns in these predictions using a soap film smoother. This work demonstrates a flexible and interpretable approach to modelling phonetic accent variation in speech recordings.

The talk will be based on joint work Davide Pigoli, John Aston and John Coleman (https://doi.org/10.1080/01621459.2019.1607357) and join work with Aranya Koshy (https://doi.org/10.1111/rssc.12555)