Sociolinguistics and forensic speech science: knowledge- and data-sharing

Vincent Hughes[1], Jessica Wormald[1,2], Tyler Kendall[3], Yvan Rose[4] and Natalie Schilling[5]

[1]*Department of Language and Linguistics, University of York, U.K.*
[2]*J P French Associates, York, U.K.*
[3]*Department of Linguistics, University of Oregon, USA.*
[4]*Department of Linguistics, Memorial University, Canada.*
[5]*Department of Linguistics, Georgetown University, USA.*

Aims

This interdisciplinary workshop will explore the benefits of closer collaboration between (socio)linguistics and forensic speech science. Specifically, we will:
- give a **practical introduction** to the application of phonetic and sociolinguistic methods in forensic casework
- discuss **data sharing** and the creation of **resources for storing and analysing recordings** for use in sociolinguistics and in forensic cases
- explore the **reciprocal theoretical and methodological benefits** of greater collaboration between fields.

The workshop fits with the 'progress' theme of the conference, as it considers current trends towards big data, which are expanding the possibilities of variationist work. The workshop also considers the societal benefits and impact of sociolinguistic work.

Background

Forensic voice comparison involves the analysis of known and unknown voices to assess whether they belong to the same or different speakers. To do this, it is essential to know the typicality of the features within the speech community. However, there is currently a lack of available data for use as reference material in forensic cases. We, therefore, believe it is essential that forensics and sociolinguistics work more closely to generate appropriate reference data. For such collaboration to work, it is important to discuss how data are collected, stored, accessed, and analysed.

We believe that closer collaboration with forensics will also have substantial benefits for sociolinguistics. The forensic scenario allows researchers to consider within- and between-speaker variation under real-world conditions. Further, forensics places the individual at the centre. For sociolinguistics, this can improve our understanding of how the individual relates to the group and the definition of speech communities, and inform models of sound change.

Contents of the workshop

Practical (60 minutes)
Participants will be provided with two forensically realistic recordings; one of an unknown offender and the other of a known suspect. Participants will be asked to compare the speech patterns with a view to answering the question: do the recordings contain the voice of the same or different speaker(s)?

We will give an overview of how this work is conducted in the UK and USA, focussing on analytic methods and conclusion frameworks. We will also present our plans for greater collaboration between sociolinguistics and forensics.

Talks (45 minutes)
There will be three 15 minute talks:

- **Yvan Rose:** Building searchable corpora for linguistic and forensic analyses
  This paper will discuss the underlying principles and structure of large, searchable corpora of speech recordings, the analytic methods that can be integrated, and consider the value of data-sharing.

- **Tyler Kendall:** Using large corpora in sociolinguistics
  This paper will exemplify how large corpora can be used to address novel questions in sociolinguistics and how such corpora can be applied in the forensic domain.

- **Natalie Schilling:** Ethical considerations and implications for the collection of speech corpora for use in forensic casework
  This talk will discuss ethical issues associated with allowing access to sociolinguistic corpora for the purposes of forensic analyses, law enforcement, and intelligence purposes.

Discussion (15 minutes)
The remaining time will be used for questions relating to the short talks and general discussion about the issues raised.