

HSYRF 2022 - Abstracts

Workshop 1

Nicola McLelland: Grammars, dictionaries, and other normative sources in historical sociolinguistic research

This workshop will introduce participants to the range of sources that attempt to codify language (grammars, dictionaries, usage guides, language columns, online fora, etc.), and to a range of different approaches to exploiting them in historical sociolinguistic research. We will consider how we might categorize such texts, and how we can analyse them as texts, and (harder!) their possible influence on language practices.

Relevant texts may be more or less explicitly or implicitly normative, prescriptive or descriptive in intention, more or less authoritative. It is useful to consider the agency and authorship of a text; its intended audience and relationship to that audience (monologic, dialogic? User input?); what can paratextual material tell us? How can we evaluate the macro- and micro-structural features of the text; the linguistic metalanguage being used; rhetorical devices; argumentational topoi; construction of authority; etc.?

Second, how can we go about assessing the relationship between codifying texts and actual usage, and what are the obstacles? We shall consider, tentatively, examples from French, German, Dutch and English.

Workshop 2

Gert De Sutter: A gentle introduction to categorical data analysis for historical sociolinguistics

This course offers an introduction to categorical data analysis which is specifically geared towards researchers in the field of historical sociolinguistics. Categorical data analysis entails the quantitative analysis of datasets in which the dependent variable (or response variable) is nominal or ordinal in nature (i.e. variables with discrete (un)ordered values, such as the linguistic choice between *metro* and *subway*), and in which it is tested to what extent one or more independent (or explanatory) variables affect the outcome of the dependent variable. Typical examples include research designs in which the choice between 2 or more synonymous lexemes (dependent variable) is related to independent variables such as time period, genre, region, text topic, and so on; other examples include the alternation between functionally equivalent morphosyntactic constructions such as the *that* or genitive alternation in English or variable word order in Dutch. Various inferential statistical analyses are presented and discussed, with a strong focus on hypothesis testing rather than data exploration; an authentic dataset on variable word order in Dutch verbal constituents will be provided (Coussé & De Sutter 2008). The following topics will be covered:

- Univariate and bivariate CDA: Chi-square, odds ratio and relative risk.
- Multivariate CDA (1): Binary logistic regression
- Multivariate CDA (2): Generalized Linear Mixed Models
- Multivariate CDA (3): Classification and Regression Trees
- Multivariate CDA (4): Random Forests

Target audience:

PhD students or postdocs with a keen interest in using statistical analysis in their research.

Course prerequisites:

Basic knowledge of working with R

Basic knowledge of statistical terminology

Masterclass

Sali Tagliamonte: Sociolinguistic Archaeology. The study of language change in community context

In this master class, I will review the methods and practises of studying language change in vernacular speech data collected via sociolinguistic fieldwork in communities, what I refer to as “sociolinguistic archaeology”, following the insight of van der Tuuk (1971) who wrote: “every language is more or less a ruin”. Residual forms (ruins) are often left behind by linguistic change offering insight into the original state of affairs. In this regard, contemporary rural dialects offer a key source of data. Relevant additional information can be found in books, articles, dictionaries and on-line resources but much can also be discovered in family attics and local archives. Another key consideration is the exact time and place of the data, its coverage and especially its context in time and space. But a critical challenge is how does the analyst spot an ideal feature to study? One way to do this is to scour the prescriptive literature for social complaints and judgements or the dialectological record for regional patterns and unique attestations. Another is to use digital mining tools to seek out potential features in the relevant data (if possible) using well known measures such as counts, frequency or type/token ratios. With these and other strategies outlined as well as some attention to accountable analysis, I will present a case study for illustration. My goal is to teach students how to use these methods themselves so that they can experience the joy of discovering how the relics of the past can offer insight into linguistic variation and underlying change.

Bad Data Keynote Lecture

Miriam Meyerhoff: Vestiges of variation and rates of change

Pawley’s Puzzle (Pawley 1981) can be stated as follows: why is there so much language diversity in Melanesia and so little in Polynesia? Part of his speculative answer to this is that Melanesian languages only recently diversified and that the pace of change picked up over time.

I have been working on one of the so-called “aberrant” North Vanuatu languages, Nkep, for the last ten years, trying to better understand the historical and synchronic basis for its typological distinctiveness. I have recently been assisted in this by material in the Paradisec collection, which includes letters that attempt to document some of the variation present in Nkep in the 1920s. The data is severely limited, and was produced by people who were non-native, though fluent, speakers of Nkep. Nevertheless, I will suggest that by drawing on this historical record, and linking this to my own observations of variation in the 2010s, we might still be unable to resolve Pawley’s Puzzle, but we may be able to enrich our understanding of the recency of Melanesian language diversification.