MAX PLANCK INSTITUTE
FOR EVOLUTIONARY ANTHROPOLOGY

Project Members

- Paul Heggarty
- Jakob Runge

'Sound Comparisons': New Tools and Resources for Exploring Language Family Diversity on the Web

This project arose out of the separate but related 🔗 Language Relatedness and Divergence research theme, which has over a number of years invested a great deal of fieldwork to collect explicitly comparative recordings on regional accents, dialects and languages across both Europe and the Andes. It then further entailed many thousands of hours of phonetic analysis and transcription to create the databases from which its measures of linguistic divergence were calculated.

To make the most of this data collection and analysis, that research has always been accompanied by this *Sound Comparisons* project, to ensure that the recording and transcription data are made as available and relevant as possible not just to other researchers in linguistics, but also to the people who actually speak the language varieties concerned. Thanks to dedicated outreach funding on a previous project, this began with devising and programming 'hover to hear' websites, a user-friendly design to allow users to hear and compare instantaneously online the precise differences in pronunciation from one region to the next, across an entire language family at a time.

The first such website was created for the main language families of the Andes, and is currently still hosted at www.quechua.org.uk/Eng/Sounds/. A second project then established www.soundcomparisons.com, on Accents of English from Around the World, later extended to the entire Germanic family at www.languagesandpeoples.com/Germanic. Those original websites have become somewhat outdated, however, and were entirely reprogrammed by Jakob Runge, into a new structure, no longer of static html pages, but a real-time database lookup system. This also allowed a wide range of powerful new functions: displaying transcriptions and instantaneous sound files on a Google map view; advanced searching and filtering, of both spellings and IPA transcriptions; selecting and exporting transcriptions and audio files for users to download; and an interface to allow the entire website to be easily translated into new user languages.

The new website is launched in spring 2013, by combining the three existing databases, plus new components from a specific sub-project on the Sounds of the Andean Languages. Over the course of the rest of 2013 and 2014 the new site was further expanded with new coverage of the Slavic and Romance families. Ultimately the whole structure was made available as a template so that new websites can easily and swiftly be added for any other language family. Candidates to be targeted in the medium and long term include Basque, Turkic, Arabic, Bantu, Arawak, Celtic, and Indo-European itself.