Sound Comparisons: Documenting, Exploring and Quantifying Linguistic Divergence in Phonetics

Paul Heggarty — DLCE, Max Planck Institute for the Science of Human History, Jena

Paul.Heggarty@gmail.com

Abstract

Sound Comparisons is a framework for explicitly comparative language and dialect documentation, 'shallow but broad', to collect, explore and qualify the phonetic diversity across any given language family. So far it covers Germanic, Romance, Slavic and Celtic, three major indigenous language lineages of the Andes (Quechua, Aymara and Mapudungun), and the indigenous Oceanic language varieties of Malakula island in Vanuatu. An imperative for the project is to record the many endangered or moribund regional and dialectal varieties of these families (about 400 varieties so far).

The database consists of recordings and detailed phonetic transcriptions of a consistent sample of the same 100-200 cognate words, as pronounced right across the regional diversity in accents, dialects and languages within each family and/or geographical region covered.

This talk presents — and hopes for feedback on — the web interface to the database at www.soundcomparisons.com, now nearing completion. Based on customisable maps and tables of 'instant playback' recordings, the design aims firstly to be user-friendly for the general public, not least the speakers of these languages, to support various outreach and revitalisation objectives.

Secondly, the website also offers powerful and instantaneous query, search and filter functionality (by orthography, IPA, phonological 'wildcards', proto-forms, etc.), so that linguists can focus on any particular research question in the phonetic/phonological diversity and history of a family. One can also link to any preferred search and/or selection of language varieties and cognates, and download all corresponding sound files and transcriptions.

The project's own research objective is to quantify net divergence in phonetics — but certainly not by some off-the-shelf computational method, grossly applied to pseudo-phonetic data as abstract 'strings'. Rather, *Sound Comparisons* uses a dedicated algorithm, custom-designed for phonetics, explicitly informed by the architecture of phonetic classification, and by expert comparative and historical linguistic knowledge. Only by incorporating both can we aspire to a precise and truly meaningful expression in numbers of actual linguistic significance.