Going global: new databases, new analysis tools and the Glottobank project

Quentin D. Atkinson[1,2] and Russell D. Gray[1,2]

1. School of Psychology
University of Auckland
Private Bag 92019
Auckland
NEW ZEALAND

2. Department of Linguistic and Cultural Evolution
Max Planck Institute for the Science of Human History
Kahlaische Strasse 10
D-07745 Jena, GERMANY

Email: q.atkinson@auckland.ac.nz

A sizeable proportion of the world's languages are described in some detail, but much of this information is currently not systematically coded and organized. Large collaborative databases are revolutionizing fields as diverse as physics, biology and psychology. The burgeoning quantity and quality of readily available and systematically coded data is changing the nature of research in these fields. These methods are now also being applied to study the world's linguistic diversity. In this talk we describe the Glottobank project, a collaboration between the Department of Linguistic and Cultural Evolution at the Max Planck Institute for the Science of Human History, the School of Psychology at University of Auckland, and the Australian Research Council Centre of Excellence of the Dynamics of Language at the Australian National University. Glottobank is an international research consortium established to document and understand the world's linguistic diversity. Glottobank team members are pursuing this goal on two fronts. First, we have established four global databases documenting variation in language structure (Grambank), lexicon (Lexibank), paradigm systems (Parabank) and phonetic changes (Phonobank). In doing so, we seek to develop new methods in language documentation, compile data on the world's languages and make this data accessible and useful to linguists and other disciplines. Second, we are developing computational modelling tools to use this data to make inferences about human prehistory, relationships between languages and processes of language change. We will outline how Glottobank is working towards these goals and how outputs from Glottobank data can be synthesized with data from a number of other fields.