

Jakarta Field Station

Its Data and How it was Collected

Bradley Taylor

The Jakarta Field Station was a major field project of the Department of Linguistics at the Max Planck Institute for Evolutionary Anthropology. Based in Jakarta, with field assistants working in various locations across Indonesia, it operated between 1999 and 2015 with the primary purpose of recording and documenting languages of the region. Together with collaborating projects and scientists, it gathered around 2.3 million transcribed utterances comprising around 8.7 million word tokens from primarily naturalistic language recordings. In 2016 this data is to be openly released.

This presentation will give a brief, primarily technical, overview of the data - its size, form, relationships, selected counts - and how it was collected and organised. During the lifetime of the Field Station over one hundred people worked in different parts of the archipelago and at different times, recording, annotating, and submitting linguistic material. Having a clearly defined, workable methodology was critical and this presentation will include discussion of the tools and work processes involved, as well as some of the difficulties experienced along the way.