Trevor Johnston: *Adding Value to, and Extracting Value from, a Signed Language Corpus through Strategic Annotations*

Trevor Johnston is Professor of Sign Language Linguistics in the Department of Linguistics, Macquarie University, Sydney, Australia.

He was born in Australia and earned his Ph.D. in Linguistics at the University of Sydney in 1989.

He is Chief on two current projects, first of which is concerned with Auslan/English interpreter-mediated communication in health and medical settings focusing on medical terminology in Auslan using Auslan Signbank (an internet-based dictionary). The second project then applies comparison of variation in Auslan and BSL for evidence of the hypothesis that grammar is emergent from language use, is driven by frequency and usage which is reflected in the processes of grammaticalization and language change.

He also conducted research in the area of Auslan assessment, especially as a first language, in the evaluation of sign bilingual education programs.

He coined the term "Auslan" and authored several dictionaries among which is also the very first Auslan dictionary that was completed in 1989 and then revised with Adam Schembri and named "Signs of Australia" in 1998.

Research interests: Bilingual Deaf Education, Corpus Linguistics (both for research into sociolinguistic variation and for empirical language description), Gesture Studies, Grammatical Description, Grammaticalization, Language Evolution, Lexicography

Abstract

This lecture I explain the way in which multimedia annotation software is being used to transform an archive of Auslan recordings into a true machine-readable linguistic corpus. After the basic structure of the annotation files in the Auslan corpus is described and the exercise differentiated from transcription, then the glossing and annotation conventions are explained. Following this the searching and pattern-matching at different levels of linguistic organization these annotations make possible is exemplified. I explain why it is important to be clear about the difference between transcription and annotation otherwise at the end of the exercise—despite time consuming and expensive processing of the video recordings— we may not actually be able to discern the types of patterns in our corpora that we had hoped we could. The conventions are designed to ensure that the annotations really do enable researchers to identify regularities at different levels of linguistic organization in the corpus and thus to test, or build on, existing descriptions of the language.