

Using technologies and corpora to enhance research into Auslan, Australian indigenous sign languages, and co-speech gesture.

The presentation reports a project to secure Auslan (Australian Sign Language) resources at the Monash University node of the Language Data Commons of Australia (LDaCA). The resources are Auslan Signbank, a web-based multi-media dictionary, and the Auslan Corpus, a collection of video recordings of the language which have been annotated in ELAN annotation files (Smith, et al 2022). The Monash node will also include data from Australian indigenous sign languages, Australian Deafblind signing, and co-speech gesture (Willoughby, et al under review).

The LDaCA project aligns well with the theme of this ICL conference: "Languages, Communities, Technologies". LDaCA aims to establish a sustainable long-term repository for ingesting and curating language data collections of national significance; to democratise access to Australia's rich linguistic heritage through enabling those collections to become more FAIR while following the CARE principles; and to demonstrate how to balance research needs with preserving community rights. It also aims to develop the computational capabilities, technical infrastructure and support services to analyse language collections at scale. It has three phases: (1) securing language data collections, (2) aggregating and linking collections, and (3) enhancing their research potential and facilitating new research.

In this presentation we report on the Auslan resources in phases 1 and 2 (2022-24) and describe plans for phase 3 (from 2024/25). The aim of the Auslan node of the LDaCA project is to make Auslan resources accessible to the language community, encourage community participation in the curation of the data, and facilitate and extend their uses in language teaching and linguistic research. The software platforms of both resources are now shared with compatible with other LDaCA resources for spoken languages. The Auslan resources have been aggregated and linked with each other so that (i) users of the dictionary can view attested corpus examples for an entry; and (ii) users of the corpus can instantly view the dictionary entry for an already glossed sign to check phonological, lexical and grammatical information about it, and/or to ensure that the correct annotation gloss for a sign token has been chosen. This will enhance additions to annotations in the Auslan Corpus, entries in Auslan Signbank and the integrity of research based on both.

Reference

Smith, R. T., Willoughby, L., and Johnston, T. (2022). Integrating Auslan Resources into the Language Data Commons of Australia. *Proceedings of the 10th Workshop on the Representation and Processing of Sign Languages, Language Resources and Evaluation Conference (LREC2022)*, Marseille, 20-25 June 2022, LREC.

Willoughby, L., Smith, R. T., and Johnston, T. (under review) The GeSCA respository: Gesture and Sign Corpus of Australia. *Australian Journal of Linguistics*.