EnoLEX: A Diachronic Lexical Database for the Enggano language

Gede Primahadi Wijaya Rajeg, Daniel Krauße, Cokorda Pramartha

In an effort to document and understand the diachronic change of Enggano (Glottocode: engg1245; ISO 639-3: eno), an endangered Austronesian language spoken on Enggano island in Indonesia, our team presents EnoLEX, a diachronic lexical database comprising legacy material starting as early as 1850. The records come in different shapes and sizes. Some are simple listings of Dutch lexemes with their Enggano equivalents (Boewang 1854; Von Rosenberg 1855; Van der Straaten & Severijn 1855; Walland 1864; Helfrich & Pieters 1891), others are comparative wordlists (Oudemans 1879; Modigliani 1894; Kasim et al. 1987), dictionaries (Helfrich 1916; Kähler 1987), or texts interspersed with Enggano expressions (Helfrich 1888). Our database is complemented with fieldwork material of contemporary Enggano gathered between 2018-2024.

The curation process includes the etymologisation of word forms in a spreadsheet to ensure that historically unrelated forms are assigned individual IDs even though they are glossed the same in the original records (e.g. *kiho* 'ant' and *akiaki* 'ant' are different entries). This step will facilitate the investigation of diachronic change of the same concept without being misled by the presence of unrelated forms. The collation and etymologisation of the lists are carried out by a team of experts in comparative historical linguistics.

The second step includes retro-digitising (i) the Enggano-German dictionary (Kähler 1987) (through an online database entry system for crowdsourcing the manual typing of entries), and (ii) the Dutch-Enggano dictionary (Helfrich 1916) through OCR and manual correction. Finally, we standardise different orthographies across the wordlists using the set-up common orthography. The database is implemented using `Golang` (back-end) and `React` (front-end).

EnoLEX will constitute one of Indonesia's first publicly available, searchable diachronic databases of an endangered language. EnoLEX spans material of over 150 years allowing historical, statistical, dialectal, and comparative analyses of phonological, semantic, and lexical change within the same language. Our work contributes to the growing presence of lexical databases not only for linguistics (List et al. 2022) but also for investigating human prehistory and cultural evolution (Gray, Drummond & Greenhill 2009).

Boewang, Djoeragan. 1854. Verslag omtrent het Eiland Enggano. *Tijdschrift voor Indische Taal-, Land-en Volkenkunde* 2. 379–393.

Gray, R. D., A. J. Drummond & S. J. Greenhill. 2009. Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement. *Science*. American Association for the Advancement of Science 323(5913). 479–483. https://doi.org/10.1126/science.1166858.

Helfrich, Oscar Louis. 1888. De eilandgroep Engano. *Tijdschrift van het Koninklijk Nederlandsch Aardrijkskundig Genootschap* 5. 272–314.

Helfrich, Oscar Louis. 1916. Nadere bijdrage tot de kennis van het engganeesch. *Bijdragen tot de taal-, land- en volkenkund*. Leiden, The Netherlands: Brill 71(1). 472–555. https://doi.org/10.1163/22134379-90001719.

Helfrich, Oscar Louis & J. A. J. C. Pieters. 1891. Proeve van een Maleisch-Nederlansch-Engganeesch woordenlijst. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 34,

35. 539–623, 228–233.

Kähler, Hans. 1987. *Enggano-Deutsches Wörterbuch* (Veröffentlichungen Des Seminars Für Indonesische Und Südseesprachen Der Universität Hamburg 14). Berlin; Hamburg: Dietrich Reimer Verlag.

Kasim, Yuslina, Marjusman Maksan, Syamsir Arifin & Zailoet Zailoet. 1987. *Pemetaan bahasa daerah di Sumatra Barat dan Bengkulu*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa. https://repositori.kemdikbud.go.id/1684/. (28 January, 2024).

List, Johann-Mattis, Robert Forkel, Simon J. Greenhill, Christoph Rzymski, Johannes Englisch & Russell D. Gray. 2022. Lexibank, a public repository of standardized wordlists with computed phonological and lexical features. *Scientific Data*. Nature Publishing Group 9(1). 316. https://doi.org/10.1038/s41597-022-01432-0.

Modigliani, Elio. 1894. *L'isola delle donne*. Milano: Ulrico Hoepli. https://www.google.co.uk/books/edition/L_isola_delle_donne/gksCAAAAMAAJ?hl=en &gbpv=0.

Oudemans, Jean Abraham Chrétien. 1879. Woordenlijst van de Talen van Enggano, Mentawei en Nias. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* XXV(5–6). 484–488.

Van der Straaten, Johannes & Pieter Severijn. 1855. Verslag van een in 1854 bewerkstelligd onderzoek op het eiland Engano. *Tijdschrift voor Indische Taal-, Land-en Volkenkunde* 3. 338–69.

Von Rosenberg, Hermann. 1855. Beschrijving van Engano en van deszelfs bewoners. *Tijdschrift voor Indische Taal-, Land-en Volkenkunde* 3. 370–386.

Walland, J. 1864. Het eiland Engano. *Tijdschrift voor Indische Taal-, Land-en Volkenkunde* 14. 94–124.