Computer-Assisted Language Comparison in Practice

Tutorials on Computational Approaches to the History and Diversity of Languages

The Linguistic Questionnaire for Asia and Africa (ILCAA-LQ)

Asako Shiohara / Yasuhiro Yamakoshi
Research Institute for Languages and Cultures of Asia and Africa
Tokyo University of Foreign Studies

The Linguistic Questionnaire for Asia and Africa (ILCAA-LQ), published in 1966–67 by ILCAA, Tokyo University of Foreign Studies, is a 2,000-item basic vocabulary questionnaire translated into eleven languages, designed for fieldwork elicitation in lesser-studied languages, primarily those of Asia and Africa. This paper outlines its structure, its item selection, and its relation to earlier questionnaires and word lists such as Swadesh's list, introduces its two digitized versions, ILCAA-LQ 2000 and ILCAA-LQ 2010 Expanded Edition, and surveys the vocabularies compiled with it since its publication.

1 Introduction

The Linguistic Questionnaire for Asia and Africa (ILCAA-LQ) was published in 1966–67 by the Research Institute for Languages and Cultures of Asia and Africa (ILCAA) at Tokyo University of Foreign Studies (ILCAA 1966–67). It was developed as a tool for collecting lexical items from lesser-studied languages, primarily those of Asia and Africa. The ILCAA-LQ consists of two volumes. The first volume (1966) includes the first 1,000 Japanese items along with their English and French translations. It takes the form of a compact (105 × 175 mm) notebook, designed for researchers to fill in collected words directly during fieldwork. The second volume (1967) includes all 2,000 Japanese items with translations into eleven languages: Arabic, Chinese, English, French, German, Hindi, Indonesian (Malay), Persian, Russian, Spanish, and Swahili. Thus, the ILCAA-LQ can function as a comparative conceptual dictionary.

Computer-Assisted Language Comparison in Practice
Volume 9, Number 2
URL: https://calc.hypotheses.org/9678 DOI: 10.15475/calcip.2026.2.6
Published under a Creative Commons Attributions 4.0 LICENSE
Published on 17/08/2026

2 Structure of the ILCAA-LQ

The 2,000 items included in the ILCAA-LQ were classified into four classes: A (200 items), B (300 items), C (500 items), and D (1,000 items). The items were selected from concepts included in various publications, including three linguistic questionnaires (Hattori 1957; Cohen 1950–51; École française d'Extrême-Orient 1938), 17 basic vocabulary lists from various languages, and two Japanese dictionaries. As the first step, four researchers selected approximately 60,000 items from these resources; the 2,000 items were then selected from among them and classified by the seven members of the editorial committee.

As stated in the ILCAA-LQ's introductory essay, "How the Language Questionnaire Was Made," the A class shares 115 items with M. Swadesh's 216-item list for glottochronology. Although the essay gives no reference, this most likely refers to the 215-item list first proposed in Swadesh (1950). The authors of the essay regard this as a fairly close match, while noting that closer agreement might have been expected. They attribute the discrepancy to differences in the linguistic basis of the two lists: Swadesh's list is biased toward English and American Indian languages, whereas the ILCAA-LQ is biased toward Japanese and Asian and African languages. As an example, they point out that the ILCAA-LQ distinguishes four kinship terms—兄 "older brother," 姉 "older sister," 弟 "younger brother," and 妹 "younger sister"—reflecting relative age distinctions that are lexically unmarked in English. Swadesh's (1950) list, for its part, includes only "brother (elder)" and "sister (elder)."

3 Digitized ILCAA-LQ (ILCAA-LQ 2000 & ILCAA-LQ 2010 Expanded Edition)

The ILCAA-LQ has been digitized in two versions, ILCAA-LQ 2000 (Minegishi & Hayasi 2000) and ILCAA-LQ 2010 Expanded Edition (GICAS Project 2010), both in Excel format. ILCAA-LQ 2000 was developed on the initiative of two ILCAA researchers, Makoto Minegishi and Tooru Hayasi. It contains the item ID, Japanese (in standard orthography and in kana transcription), and an English translation, along with correspondences to Hattori (1957), another questionnaire widely used by Japanese researchers for lexical elicitation. The Hattori (1957) columns comprise the Hattori classification number, ID, Japanese (orthography), Japanese (kana transcription), and English. ILCAA-LQ 2010 Expanded Edition was developed by the GICAS Project, led by Kazuhiko Machida at ILCAA. It contains the item ID, Japanese (orthography and kana transcription), translations into all eleven languages included in the printed version, and a Tamil translation (Tamil script and romanization) prepared by Norihiko Uchida. The two versions were first published on the ILCAA website in 2000 and

2010, respectively, and are now available, alongside a PDF of the original printed ILCAA-LQ, through Prometheus-Academic Collections, the TUFS institutional repository, under a CC BY 4.0 license (DOI: 10.15026/0002001820).

4 Impact of the ILCAA-LQ

Since its publication, the ILCAA-LQ has been widely used for lexical elicitation in fieldwork on lesser-studied languages. Its 2,000-item inventory has provided a common reference for vocabularies of a range of languages (e.g., Wada 1973; Hayasi 1996; Minegishi & Murmu 2001; Tangiku 2008; Shirai 2011; Shinagawa et al. 2024). More recently, the same inventory has also been used to index online lexical databases and datasets, including Yamakoshi (2017–2025) and Shiohara et al. (2024), facilitating comparison across independently collected lexical resources. Once the ILCAA-LQ is added to Concepticon, these resources can be linked to the many other datasets mapped to Concepticon concept sets.

5 Conclusion

As a list of 2,000 basic concepts, the ILCAA-LQ has been used as a tool for collecting lexical items from lesser-studied languages. This inventory has provided a common reference for vocabularies of a range of languages. The ILCAA-LQ also has the potential to serve as a standard questionnaire for future lexical research. For the future, its integration into the Concepticon reference catalog (https://concepticon.clld.org, List et al. 2026) could further enhance this potential by allowing lexical resources based on the ILCAA-LQ to be linked to datasets built from other questionnaires.

References

Cohen, Marcel. 1950–51. *Questionnaire linguistique*. Nimègue [Nijmegen]. 570pp.

École française d'Extrême-Orient. 1938. *Questionnaire linguistique* [sic: "liguistique" in original]. Hanoi. 491pp.

GICAS Project. 2010. *ILCAA-LQ 2010 Expanded Edition* [Dataset]. Fuchu: ILCAA. https://doi.org/10.15026/0002001820 (file: Linguistic_Questionnaire_2010.xls)

Hattori, Shirō (ed.). 1957. *Kiso goi chōsahyō* [Basic vocabulary survey sheet]. Tokyo: Department of Linguistics, Faculty of Letters, University of Tokyo.

Hayasi, Tooru. 1996. *Gendai Uiguru-go Urumuchi hōgen goishū* [Vocabulary of the Urumqi dialect of Modern Uyghur] (Asian and African Lexicon Series 31). Tokyo: Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies. 271pp.

ILCAA (Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies). 1966–67. *Ajia Afurika gengo chōsahyō* [Linguistic questionnaire for Asia and Africa]. 2 vols. Tokyo: ILCAA.

List, Johann Mattis, Annika Tjuka, Frederic Blum, Alžběta Kučerová, Carlos Barrientos, Christoph Rzymski, Simon Greenhill, and Robert Forkel (eds.) 2026. CLLD Concepticon [Dataset, Version 3.4.1]. Zenodo. https://doi.org/10.5281/zenodo.21373838

Minegishi, Makoto & Tooru Hayasi. 2000. *ILCAA-LQ 2000* [Dataset]. Tokyo: ILCAA. https://doi.org/10.15026/0002001820 (file: Linguistic_Questionnaire_2000.xls)

Minegishi, Makoto & Ganesh Murmu. 2001. *Santali basic lexicon with grammatical notes* (Asian and African Lexicon Series, No. 38). Tokyo: Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies. 268pp. https://doi.org/10.15026/0002000648

Shinagawa, Daisuke, Ryota Ueno, Tsubasa Takahashi, Toshio Tôyama, Sakura Ishikawa & Ayaka Yamamoto. 2024. *Copperbelt Bemba: A collection of phonetic materials on nominal tone patterns* (ILCAA Intensive Language Course 2023 “Bemba” Textbook 2). Fuchu: Research Institute for Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies. 167pp. https://doi.org/10.15026/0002000594

Shiohara, Asako, Yanti & Dominikus Tauk. 2024. *Bolok Helong lexical data* [Dataset]. https://doi.org/10.15026/0002000458

Shirai, Satoko. 2011. *Sensei minzoku sōrō no gengo tokuchō ni kansuru kisoteki hōkoku: Dapa-go to no taishō o chūshin ni* [A basic report on the linguistic features of the Sichuan ethnic corridor: With a focus on comparison with Dapa]. Nagoya: Nagoya Institute of Technology. 88pp.

Swadesh, Morris. 1950. Salish internal relationships. *International Journal of American Linguistics* 16(4). 157–167.

Tangiku, Itsuji (ed.). 2008. *Basic vocabulary of the Sakhalin dialect of Nivkh language: Nogliki dialect* (Asian and African Lexicon Series 51). Fuchu: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

Wada, Shohei. 1973. *Iraku-go kiso goishū: Suwahiri-go taishō* [Iraqw basic vocabulary with Swahili equivalents]. Tokyo: Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

Yamakoshi, Yasuhiro. 2017–2025. *Database of basic vocabulary in Mongolic languages* [Dataset]. https://mongolicbv.aa-ken.jp

Supplementary Data
The data mentioned in this study is available through an open repository with a permissive license (see https://doi.org/10.15026/0002001820).