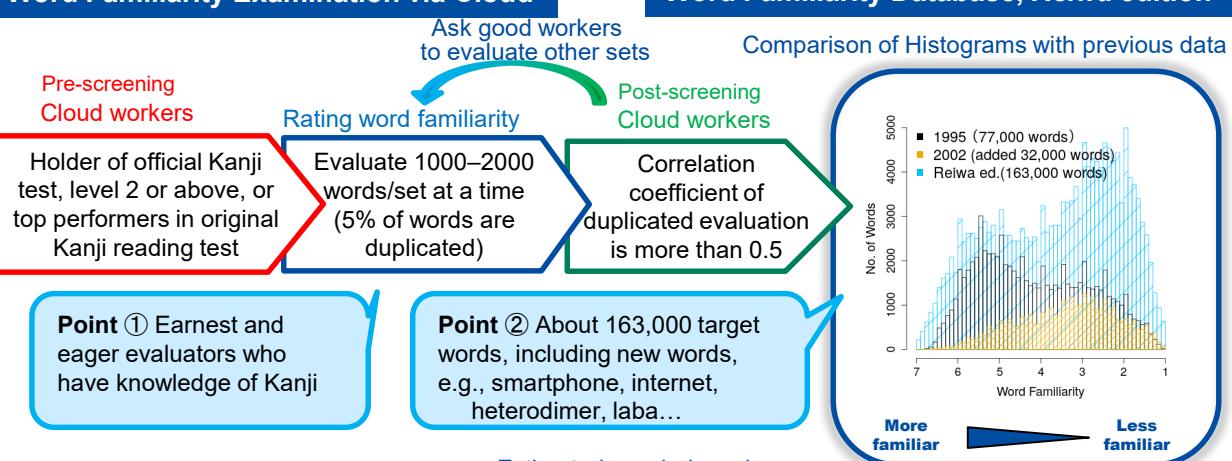


Abstract

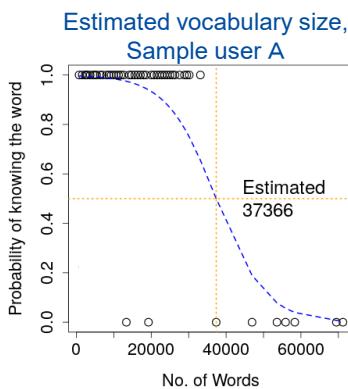
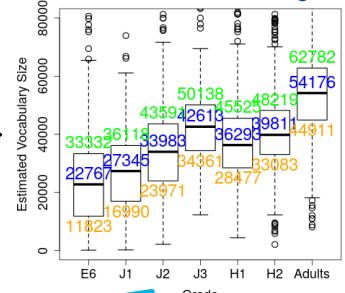
More than 20 years ago, NTT conducted psychological experiments to investigate word familiarity and thus construct a Japanese lexicon of about 77,000 words. This lexicon is still used in many fields.

Now, we reinvestigated and reconstructed an unparalleled-scale Lexicon of 163,000 words through crowdsourcing. By applying careful screening, we succeeded in obtaining highly reliable results. This makes it possible to make comparisons with results of 20 years ago. Based on this, we created the Reiwa edition of a vocabulary-size estimation test. Furthermore, this lexicon allows us to estimate the vocabulary size appropriate for the present day. We are working on vocabulary-size investigations for a wide range of ages, including elementary school to high school students. In the future, we will investigate and analyze the relationship between vocabulary size, reading comprehension, and academic ability, aiming to achieve effective educational support.

Word Familiarity Examination via Cloud**Word Familiarity Database, Reiwa edition****Vocabulary-size estimation**

Please select words that you know.

- bank
- charge
- ...
- intimacy
- recast

**Estimated results for each grade**

Point ③ Checklist of only about 50 words

Vocabulary-size test:
<http://www.kecl.ntt.co.jp/icl/lirg/resources/goitokusei/>



Point ④ About 4,600 persons, including elementary/junior high/high school students

References

- [1] S. Fujita, T. Kobayashi, "Reexamination of word familiarity and comparison with past examination," *The 26th Annual Meeting of the Association for Natural Language Processing*, pp. 1037-1040, 2020.
- [2] S. Fujita, T. Kobayashi, T. Yamada, S. Sugawara, T. Arai, N. Arai, "Vocabulary size of elementary, junior high and high school students and analysis of relationship with word familiarity," *The 26th Annual Meeting of the Association for Natural Language Processing*, pp. 355-358, 2020.

Contact

Sanae Fujita, email: cs-openhouse-ml@hco.ntt.co.jp
Linguistic Intelligence Research Group, Innovative Communication Laboratory

