

Frequency-based word lists provide good indications of which words language learners should learn, but are less useful in revealing which words learners actually know

What this research was about and why it is important

Learning vocabulary is one of the most important aspects of learning any foreign language, and a significant part of any language class involves vocabulary instruction. The design of dictionaries, language curricula, teaching materials, and tests is often informed by word lists that show how frequent words are in contexts where that language is dominant (e.g., U.S. or U.K. for English), as this provides an estimation of how *useful* a word might be. An ideal teaching sequence will: 1) prioritize frequent/useful words over those that are rare/obscure, and at the same time 2) focus only on words that the learners do not already know, to optimize class time. While the first criterion might be reasonably met using word frequency lists, the second criterion is more difficult to achieve, and often word-frequency lists have been the *de facto* method by which learner word knowledge is estimated, despite the fact that words are not learned in the same order by learners from different foreign language learning contexts. This study tested the ability of several widely-used English word-frequency lists to predict word knowledge among Japanese university students and found that word knowledge is only predicted well by specific kinds of word-frequency lists, and very poorly by others. The authors also found that the common practice of treating related words that share the same root as the *same word* in word lists may not be appropriate for English language pedagogy in Japanese university settings.

What the researchers did

- Word frequency lists were generated from 18 types of English text collections (corpora). These ranged in type (fiction, newspapers, academic, magazines, conversation and formal speech), dialect (UK vs. US) and modality (spoken vs. written). The size of these corpora ranged from 4 million to 12 billion words.
- 625 students at 4 Japanese universities took a 149-word vocabulary test in which word difficulty ranged greatly.
- For each tested word, the degree that the word was known by the group of participants (i.e., % known) was compared to how frequent the word is in each of the 18 different English corpora. The correlations between word knowledge and word frequency were then compared to reveal which kinds of word frequency lists were good predictors of Japanese student word knowledge.
- Additionally, the authors compared word frequency lists where the definition of *word* was altered: 1) Word family-based lists assume that words that share the same root (*nation*, *nations*, *national*, *nationalization*, etc.) are the same word, 2) inflection-based lists, which do not make the assumption that learners understand all the members of a word family if the root word is known (i.e., only *nation* and *nations* are assumed to be the same word)

What the researchers found

- Word lists widely used in foreign teaching settings were poor predictors of word knowledge of Japanese students.
- Word lists generated from texts of children's TV, conversational speech, and fiction were the best predictors.
- Word lists that were based on word-families did not predict word knowledge as well as inflection-based word lists.

Things to consider

- Foreign language instruction should include English word formation (prefixes, suffixes, etc.) as a learning objective, rather than assuming learners are able to infer the meanings of words that share the same root.
- Language teachers and language materials developers struggle to estimate the difficulty of words in texts that they are considering using in class without trial and error. Furthermore, it is difficult to know precisely which words in these texts are already known by learners, and which should be glossed, pre-taught, included in course learning objectives, and/or replaced with less-obscure synonyms. Use of improved word lists can assist with these decisions.
- Word lists that more accurately estimate the likely sequence in which speakers of different first languages (e.g., Chinese, Arabic, Japanese, German, Spanish, Vietnamese, etc.) learn English words are currently being developed, and these lists will ultimately lead to better diagnostic vocabulary tests and readability formulae for specific groups of learners. These custom tools will allow learners to be more accurately matched to level-appropriate texts, and for syllabi to include more appropriate vocabulary learning objectives: useful words that are not already known.

Material, data, open access article: N/A.

How to cite this summary: Pinchbeck, G.G., Brown, D., Kramer, B. (2022). Frequency-based word lists provide good indications of which words language learners should learn, but are less useful in revealing which words learners actually know. *OASIS Summary* of Pinchbeck, Brown, McLean, et. al. (2022) in *System*. <https://oasis-database.org>

This summary has a [CC BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license.