Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. List of Affixes in the NFL7.
Appendix S2. Sublists of the NFL7.

Appendix: Accessible Summary (also publicly available at https://oasis-database.org)

A New Type of Word List to Reduce the Learning Burden

What This Research Was About and Why It Is Important

Frequency lists are an important tool in vocabulary research and instruction, especially as used in text profiling computer software. They make it possible to sequence and plan the lexical component of language learning, for example, analyzing texts in terms of lexical difficulty. But there is an issue about how these lists should be structured. Words in lists clearly have to be "grouped," otherwise the lists would be enormous with many of their items nearly identical. The main groupings are word lemmas (with just inflected, i.e., grammatical forms, like the verb *hunt, hunts, hunting*, and *hunted*) and word families (those, plus other related words of different parts of speech, called derived forms, like *hunter* and *huntress*). Each has strengths and weaknesses in a pedagogical context. Lemmas are normally recognizable as forms of the same word, but their use entails impractically large lists of similar but separated items (*hunting* and *hunter*); families produce lists with fewer units, but containing many items of widely different frequencies that learners may not recognize as related (*huntress*). To resolve this dilemma, we built a list which, while family based, contains only the most frequent family members, whether inflected or derived; that is, the nucleus of each family. We describe a way of building such a list and we compare its "text coverage" against that given by other published lists (both family and lemma based); that is, we checked how many words from each type of list are found in different texts. We describe our method of balancing list size against text coverage. We also analyze the derived forms in our list for teachable patterns. The Nuclear Frequency List (NFL) has a similar coverage to lists more than double its size; its derived forms employ a small number of morphological patterns; and it thus resolves the "grouping dilemma."

What the Researchers Did

The researchers first built a computer program that produces the NFL.

- The NFL reduces a complete family list to just the family members that are frequently used. The complete list they reduced uses the frequency information from large digital collections of words taken from real usage: The British National Corpus (BNC) and the Corpus of Contemporary American English (COCA).
- The researchers described the decisions that need to be taken when reducing a complete list to its nucleus.
- They evaluated the NFL in terms of the coverage it provides of the types of texts learners are likely to encounter.
- The researchers further analyzed the derived-word component of the NFL, in order to determine whether there exists a core of frequent affixes.

What the Researchers Found

- The 3,000 word families of the NFL contain 7,293 individual words and give an average coverage of 84.5% over a range of learner-oriented corpora, compared to the same 3,000 families in the BNC/COCA (from which it is derived) that contain 19,062 words and give a coverage of 90%.
- More than 85% of the derived forms in the NFL employ just 22 affixation patterns.

Things to Consider

- In cost–benefit terms, the loss of 5.5% coverage for about 11,000 fewer words to learn is a good trade.
- 22 affixation patterns are learnable, compared to the 100s of a complete family list.
- The NFL provides the nucleus of a lexicon that can serve both productive and receptive knowledge.

Materials, data, open access article: The Nuclear Family List-builder is available at https://www.lextutor.ca/freq/nuclear/, and the list itself is available at iris-database.org

How to cite this summary: Cobb, T., & Laufer, B. (2021). A new type of word list to reduce the learning burden. *OASIS Summary* of Cobb & Laufer (2021) in *Language Learning*. https://oasis-database.org