

## Notes from Vocabulary Research (Lewis)

Couple of sources

Learning Vocabulary in Another Language (Nation, Cambridge University Press, 2001)

Vocabulary Size, Text Coverage, and Word Lists (Nation, Waring)

### How much vocabulary do learners need to know?

- Number of words in a language
- Number of words known by native speakers
- Number of words needed to use the language

### How many words are there in English?

- Different ways of defining “word”
  - Trying to do this since the early 1900s, but within past 10 years, some agreement on how to count words
- Dictionaries
  - Webster’s 3<sup>rd</sup> New International (largest non-historical) (Dupuy, 1974; Goulden, Nation, Read 1990)
    - 267,000 entries
    - 114,000 word families excluding proper nouns
  - Same dictionary, different count
    - 54,000 word families (in Nation and Waring)
- 4 ways to count words
  - Tokens, running words
    - Every occurrence of a word
  - Types
    - Every different word
  - Lemmas (learning burden is considered)
    - Headword, inflected forms (plural, tenses, comparative, superlative, possessive), reduced forms (contractions); same part of speech.
    - Brown corpus, 1982 (no comparatives, superlatives, combined parts of speech, e.g., walk (n), walk (v))
      - 61,805 types become 37,617 lemmas
  - Word families
    - Base word, inflected forms, small number of derived forms (police, policed, policing, policeman)

### How many words does a second language learner need to know?

- Normal adult conversation: use 1,000 most common words

- But: there are great differences between oral and written language
- Well-educated adults know about 20,000 word families
  - English learners: need to read/understand at least 3,000 word families to handle academic texts (80% comprehension level)---still will need vocabulary support until 95% level is reached.
  - Students abroad are advised to know 10,000 words before going to American universities.
- Rate of acquisition
  - Native speakers learn about 1,000 word families a year up through their twenties
  - Adult English learners may never exceed 5,000 word families (Jamieson, 1976).
  - Young English learners can match the native rate, but don't close the initial gap
    - Native English children have a receptive vocabulary of about 4,000-5,000 words when they start school.
      - Task is different: once they learn to read a word, they know what it is, but an English learner has to decode it AND learn what it means.
  - Swedish student study
    - Immersed in Swedish, learning English as a foreign language
    - After 4 years, they knew 1500 words (not sure if these are tokens or word families)
    - After 6 years: 3500 words
    - After 9 years: 7000 words
- Basic Global English (GBE)
  - Form of English for businessmen to use: 850 words, 250 additional words, simple rules.
- Three Little Pigs (500 words)
  - High frequency words
    - 124 types (different words)
    - Most frequent: the, little, pig, house, a, and, said, he, I, me
- Article about forestry
  - General Service List's 2000 words: 80% of text
    - Function words (in, the, for)
    - Content words (government, forests, represent, boundary)
  - Academic words (policy, adjusted, sustained)
  - Technical words (indigenous, podocarp, beech)
    - About 5% of running words
  - Low frequency words (zoned, perpetuity, pastoral)
    - 5% of running words in academic text
      - Proper nouns

- Words almost in the list of 2000 words
- Economics text
  - 295,294 words, 5,438 word families (Sutarsyan, Nation, Kennedy, 1994)
  - 1<sup>st</sup> 1000 word families (GSL): 71.4%
  - 2<sup>nd</sup> 1000 word families (GSL): 4.7%
  - Academic Word List (570 words): 10.0%
  - Other vocab: 13.9%
- Academic corpus (Coxhead, 1998)—science, arts, commerce, law texts totaling 3.5 million running words
  - 2000 word families: 80%
  - Academic, technical, low-frequency vocab: 20%
- Texts totaling 5 million running words (Carroll et al., 1971)
  - Number of words: 86,741
  - 3,000 word list: 85.2%
  - 2,000 word list: 81.3%
  - 1,000 word list: 74.1%
  - 100 word list: 49%
- Brown corpus (various 2,000 word texts of American English totaling 1 million running words)
  - 1000 lemmas: 72%
  - 2000 lemmas: 79.7%
  - 3000 lemmas: 84%
  - 5000 lemmas: 88.6%
- American Heritage Word Frequency Book (Carroll et al., 1971)
  - Printed school English: 88,500 word families (Nagy, Anderson, 1984)
- General Service List (Michael West)
  - 2000 word families
  - 165 word families are function words (a, some, two, because, to)
  - Others are content words (nouns, verbs, adjectives, adverbs)
  - Older graded readers are based on this list
  - About 80% agreement between well-designed high frequency lists
  - 1<sup>st</sup> 1000 cover about 77% of academic text, and the 2<sup>nd</sup> 1000 about 5%
  - Range: word families occurs in many different kinds of texts
  - These words are the most important for ELD students to learn
- Academic Word List (University Word List)
  - 570 words that are not in the basic 2000, but occur over a wide range of academic texts.

- Adding this to the 2000 words, coverage of academic texts goes from 78% to 87%.
- Better to do this list before the 3<sup>rd</sup> 1000 words
- Assume, establish, indicate, conclude, maintain, analyze, assess, concept, definition, establish, categories, seek)
- Specialized or technical vocabulary
  - About 10% of academic texts
  - Some are on the 2000 word list or the AWL, but are used in technical ways (wall, (cell) wall)
  - Best to learn these words while studying the discipline

### Ways of teaching and learning high frequency words

- Direct teaching
  - Teacher explanation
    - Give *meaning* of the word by
      - Use L1 translation
      - Use a known English synonym
      - Show object or picture
      - Give quick demonstration
      - Break the word into parts and give the meaning of the parts and the whole word (prefix, root, suffix)
      - Give example sentences with the word in context to show meaning
      - Exercises
        - Word and meaning matching
        - Labeling
        - Sentence completion
        - Crossword puzzles
        - Categorization (by meaning, by function); semantic maps, syntactic groups)
        - Definition Diagonals (four clues to meaning)
          - [www.idra.org/lessons/definitiondiagonals.htm](http://www.idra.org/lessons/definitiondiagonals.htm)
  - Draw attention to the *form* of the word by
    - Show how the spelling is like the spelling of known words
    - Give stress pattern and pronunciation
    - Write the word on the board
    - Point out any spelling irregularity
    - Have student read the word
    - Use dictation to see if student can write the word if they hear it
    - Exercises

- Following spelling rules
- Recognizing word parts
- Building word family tables
- Draw attention to the *use* of the word by
  - Show the grammatical pattern the word fits into
  - Give a few similar common associations, or collocations (e.g., for “friend”, close, family, good)
  - Mention any restrictions (formal, impolite, only with children, old-fashioned, technical, etc)
  - Give a well-known opposite
  - Exercises
    - Sentence completion
    - Collocation matching
    - Collocation tables
    - Dictionary word
- Peer teaching
- Direct learning
  - Study from word cards
    - Old-fashioned, but Nation says efficient, effective (2001)
    - Translation on back
  - Dictionary use
  - Keyword method (Baumann & Kameenui, 1991): use a mental image (e.g., “carlin”—visualize an old woman driving a car—carlin means old woman); helps with later recall.
  - Computer-assisted learning
  - Small sets (sets of 10 words, with cumulative review)
- Incidental learning
  - Guessing from context in extensive reading
    - Need to know about 95% of words: 19 of 20 words.
    - Lexile levels: 75% comprehension rate
      - MetaMetrics, NIH grants
      - Word frequency, sentence length
  - Graded reading
  - Communication activities (listening, speaking, being understood)

### **Testing vocabulary knowledge**

- Vocabulary Levels Test
  - Receptive
    - Matching (6 words, 3 fill-in phrases)
  - Productive version (Laufer & Nation, 1999)
    - <http://www.er.uqam.ca/nobel/r21270/levels/>
    - Fill in missing word in a sentence, three or four first letters given

