CHAPTER 4

Which Words and Word Meanings Should We Teach and Assess?

At this point, the complexities of vocabulary assessment should be very clear. Vocabulary knowledge involves much more than definitions; the size of the vocabulary makes it impossible to sample more than a small subset of the words in the language, either for explicit, focused teaching or for assessment. We are afloat on a sea of words and can cast only a small net into that sea. How, then, can we trust what our net brings back? How can we know whether we are teaching and assessing the right targets?

Selecting words for teaching and selecting words for assessment are related but not identical issues. The teacher has limited time to make a difference in students' learning, so he or she must select instructional targets that will make the largest difference for the largest number of students. The assessment designer has limited time to measure differences among students and so must select assessment targets that most reliably differentiate between students at different levels of performance. When selecting words for teaching, for instance, many authors focus on root words, assuming that morphologically derived words will be easily learned (Bauer & Nation, 1993). Someone who knows the word *sleep* should have little trouble inferring the meaning of *sleepless* and *sleeplessness*. If someone knows the word *generate,* we might equally assume he or she should have little trouble inferring the meanings of *generation, generative, regenerate, regeneration,* and so forth, but there is a danger in that assumption. *Regenerate* does not normally mean "generate again," but something more specific—to grow back something that has been lost. *Generation* can mean the process of "generating something,"

but it can also refer to groups of individuals born at about the same time. Almost any process can be *generative,* but no one would refer to a power plant as "generative," no matter how efficiently it functions to generate electricity. Every word has its idiosyncrasies, even if it is derived transparently from a known root word, and learning about such idiosyncrasies is central to vocabulary acquisition. As a result, it would be a mistake to focus assessment on root words alone, no matter how appropriate it may be for teachers to adopt root-word-based pedagogical strategies.

The literature thus far has focused primarily on pedagogical motivations for word selection and so that is where we start. We return to the issue of word selection specifically for assessment later in this chapter, when we will be better placed to evaluate differences among priorities.

SELECTING WORDS, WORD PATTERNS, AND CONTEXTS FOR INSTRUCTION

Because adequate vocabulary knowledge is such a strong prerequisite for comprehension, a major principle for selecting words to teach and test is text coverage; that is, teaching words that account for a large proportion of the language in the texts students read will minimize the amount of unfamiliar words that students encounter, but how many words do students need to know to understand the texts they have to read in school? Conversely, how much should students' reading expose them to words they do not know? Some estimates suggest that students need to know at least 95% of the words in a passage in order to read it with full comprehension, although the exact conditions under which vocabulary knowledge is necessary for comprehension have not been investigated in depth (Nation, 2001).

Prioritizing Words by Frequency and Age of Acquisition

Word frequency and age of acquisition are two measures that are commonly used to identify when words are appropriate for students to encounter or to know. A word's frequency is an estimate of how common it is in the language, as measured by how often it appears in some large corpus of text. The age of acquisition is the estimated age when a learner acquires specific words. When we explore the selection of words to teach based on word frequency and the age of acquisition, we are confronted almost immediately with the much-skewed nature of word frequency distributions (Zipf, 1935, 1945). What that means is that a small set of highly frequent words appears everywhere, while most words are much rarer. In fact, just over 100 English words account for

nearly 50% of any text (Adams, 1990). However, there are many much rarer words that do most of the heavy lifting because they carry most of the semantic load of the sentences in which they appear. For example, in the previous sentence, the words *rarer, semantic,* and *sentence* are by far the least-frequent words, but without them, the sentence means almost nothing. Many of the most frequent words are referred to as *closed-class* words because there is a limited number in each grammatical category. Closed-class words include articles (*the, an*), conjunctions (*and, or, but*), and prepositions (*in, of, for, with*). These words are akin to the mortar that holds bricks—the more semantically informative words—together.

Word frequency can be taken as a fairly good proxy for the number of opportunities people have to acquire a word because the more often a word appears, the more chances people have to learn it. As a result, much of the work on selecting words for instruction has focused on word frequency. This effort can be traced back to the pioneering work of Edward Thorndike (1921) to quantify vocabulary and prioritize words for education. Thorndike identified the 10,000 words that appeared most frequently in a collection of children's books and classic literature. Later, with the help of Irving Lorge, the list was expanded to approximately 30,000 words (Thorndike & Lorge, 1944). Kučera and Francis (1967) produced word frequencies from a corpus selected from largely adult, academic texts that contained about one million words of running text. The expanding availability of ever-larger electronic text collections (corpora) has led to more precise word frequency estimates. The CELEX corpus (Baayen, Piepenbrock, & van Rijn, 1993) and TASA corpus (Zeno et al., 1995) were based on corpora containing between 17 and 18 million words. The British National Corpus was based on a corpus of 88 million words, while the HAL Corpus (Burgess & Livesay, 1998) contained about 130 million words. More recent corpora, such as the English Gigaword corpus, derived from newspaper text (Parker, Graff, Kong, Chen, & Maeda, 2011), are based on much larger corpora, containing more than 4 billion words.

Frequency is most often used to identify words that should be a priority for instruction. For instance, Hiebert (2011) classifies words into eight frequency-based zones, based on their frequency in the TASA corpus (Zeno et al., 1995). Zones 0–2 cover words that appear 100 or more times per million words of text and correspond roughly to words that should be known by students in kindergarten through second grade. Zones 3 and 4 cover words that appear from 10 to 99 times per million words and are associated with third and fourth grade. Zones 0–4 contain approximately 80% of the words in the TASA corpus. Zones 5 and beyond are not tied to any particular grade, but occur less frequently than 10 times per million words of text and include specialized

vocabulary in content areas. Zone 5 covers words that appear two to nine times per million words; Zone 6, words that appear once per million words; and Zones 7 and 8, words that appear less than once per million. Words in Zones 6–8 are least likely to be known, as they appear very infrequently in grade-level texts. Hiebert argues that because these words are so infrequent, focusing instruction on them is less effective than one might desire. We need to be judicious in the words we select from this group.

An issue to consider in using frequency to guide word selection is that words with the same frequency can play very different roles in the language. For example, the words *respected* and *ski* have similar frequencies (Gardner & Davies, 2014), yet the frequency of *respected* reflects its use across a broad variety of contexts, while the frequency of *ski* reflects its role in contexts particular to sports and recreation. To account for these differences, the field has developed measures of dispersion—the extent to which words appear in a variety of contexts across a wide range of subject matter (Carroll, 1970, 1972). Measures of dispersion are related to ideas about balance and representativeness. Dispersion helps to identify words that may be more important than their frequency suggests. Dispersion is reported along with frequency in contemporary guides to word frequency.

Several approaches to word selection consider both frequency and dispersion to sort words into categories that have different educational priorities. We have already discussed one such system in previous chapters: Beck et al.'s (2002) distinction between Tier One words (high-frequency, high-dispersion words that form part of the core English oral vocabulary), Tier Two words (somewhat lower-frequency words with high dispersion, typically found in academic texts), and Tier Three words (rarer words with low dispersion, typically domain-specific in nature). Nation (2001) makes similar distinctions among types of words, although he divides the vocabulary into four sections: high-frequency general-purpose words, academic vocabulary, low-frequency general-purpose vocabulary, and technical vocabulary. Specific word lists reviewed in earlier chapters, such as the General Service List (West, 1953), the Academic Word List (Coxhead, 2000), and the Academic Vocabulary List (Gardner & Davies, 2014) focus on one or another of these subsets.

High-frequency general words such as those on the General Service List are of particular importance in second-language teaching since they provide the foundation upon which the knowledge of rarer words must build. Developing knowledge of academic vocabulary—mid-frequency, high-dispersion words preferentially appearing in academic written texts—is particularly important for K–12 vocabulary development and

for advanced language learners. Tier Three technical vocabulary is often treated as a lower priority for vocabulary instruction since it is closely linked with specific domain knowledge and is typically taught explicitly when domain-specific content is introduced.

Many approaches to vocabulary instruction focus on targeted words selected from academic vocabulary lists as the most productive target for focused instruction (e.g., Lesaux, Kieffer, Faller, & Kelley, 2010; McKeown, Crosson, Beck, Sandora, & Artz, 2012; Snow, Lawrence, & White, 2009). Lists such as the Academic Word List (Coxhead, 2000) and the Academic Vocabulary List (Gardner & Davies, 2014) are compiled from expository text corpora. Thus an important category of words is often left out of these discussions: sophisticated words that are typically found in novels and literary texts. These words, such as *grim* or *desolate,* are not typically introduced in academic subjects such as science or mathematics. Although there is a great deal of overlap between literary and expository academic texts, the set of words most frequently encountered in narrative reading is quite distinct (Gardner, 2004; Hiebert & Cervetti, 2011). Vocabulary from the literary dimension of academic language emphasizes common human experiences and situations and tends to be descriptive or colorful. Literary words are often more imageable (e.g., *remorse, solace, surreptitious*) than words that typify expository discourse (e.g., *coordinate, variable, attribute,* and *duration*). Literary words occur on a gradation of meaning that emphasizes particular elements of a concept. For example, word pairs such as *grim/unpleasant, desolate/lonely,* and *sweltering/warm* are associated in meaning but not really synonyms. Gardner (2004) found that literary words are less likely to represent complex concepts than words typical of expository text. This characteristic is similar to Graves and Prenn's (1986) differentiation between words for which students already possess a general concept (*indigenous, embroiled, ubiquitous*) and those that require building new conceptual knowledge (*photosynthesis, circumference, government*). This is another important distinction that can be useful for selecting words for both teaching and testing vocabulary.

Frequency is closely related to another concept that is used to identify appropriate words to teach and assess: the age of acquisition. The majority of age-of-acquisition studies have focused on subjective adult estimates of the age at which they acquired a word—that is, simply asking adults when they thought they learned specific words. Empirical studies of the age of acquisition have also been conducted by ascertaining whether children of various ages had knowledge of specific words. Such studies yielded results comparable to the subjective adult studies; however, they either examined only very young children or included only a very small set of words (Dale & Fenson, 1996; Goodman, Dale, & Li, 2008; Morrison & Ellis, 2000; Morrison, Chappell, & Ellis, 1997).

Carroll and White (1973) proposed a subjective measure for estimating the age of acquisition and demonstrated that age-of-acquisition estimates predicted how quickly people could access the right word to name what was shown in a picture; these estimates were more accurate at predicting reaction time than were the predictions based on word frequency. Since then, a variety of studies have indicated that age of acquisition is an important predictor of reading comprehension and other language-related skills, although frequency remains an important predictor as well (Gilhooly & Gilhooly, 1979, 1980; Gilhooly & Logie, 1980; Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012; Morrison & Ellis, 1995; Morrison, Ellis, & Quinlan, 1992; Zevin & Seidenberg, 2002). More recently, estimates of the age of acquisition have been developed from collections of school-age text using advanced statistical techniques such as latent semantic analysis ("word maturity"; cf. Landaur, Kireyev, & Panaccione, 2011).

To a significant extent, age-of-acquisition estimates overlap efforts to estimate the grade level at which words are learned. The Living Word Vocabulary (LWV; Dale & O'Rourke, 1976) provides the largest single study of the grades at which English word meanings are acquired. More than 30,000 word meanings are assigned to a grade level in the LWV, although grade-level estimates are provided only for every other grade from fourth grade through the end of college. The LWV reflects a series of studies in which items testing vocabulary knowledge were given to multiple grades. The grade-level estimate associated with each word was calculated by identifying the first grade level at which two thirds of the students gave the correct answer for the item used to test the targeted word. Biemiller and Slonim (2001), following up on LWV grade-level estimates for the lower grades, argue that vocabulary tends to be acquired in the same sequence, whether students are at the high end of achievement (acquiring a large vocabulary early), or at the low end of achievement. However, the reason that words may seem to be learned in a certain order may be an artifact of children's exposure to the words. At least in the early grades, students tend to read similar texts within a grade level and thus are likely to learn the same words in roughly the same order.

The variability of student vocabulary size presents the greatest challenge for instruction. After second grade, when most students have mastered decoding skills, many students can read words that they do not know and therefore do not understand, at which point vocabulary size can become a critical limiting factor in reading comprehension (Biemiller, 2005, 2006; Scarborough, 2005). The point is simple. In the early grades, students know more words than they can decode, and decoding is the barrier; once they master decoding and start encountering more complex texts, they may be able to decode words, but knowledge of their meanings becomes the barrier.

Because many students enter school with much smaller English vocabularies than their peers, a critical issue for vocabulary instruction is to determine what words need to be taught to bring lower-performing students up to grade level (Biemiller, 2015; Hart & Risley, 1995; Hoff, 2003; Rowe, Raudenbush, & Goldin-Meadow, 2012). The relative difference in vocabulary size tends to increase progressively as students move to higher grades in part because of a positive feedback loop: large vocabularies facilitate reading, which in turn facilitates vocabulary growth (Duff, Tomblin, & Catts, 2015; Stanovich, 1986, 2000). Biemiller thus recommends selecting words for direct instruction that are known to higher-performing students at a given grade level but not known to lower-performing students. This is because these are likely to be the next words that most students would learn anyway (making them more susceptible to instruction) and because learning those words will help bring lower-performing students up to the level of their peers.

Frequency, especially when combined with dispersion, provides very useful data about words. However, there are a number of both technical and conceptual issues with frequency that limit its usefulness and that should provide caution about what it tells us about a particular word.

Problems with Measures of Frequency and Age of Acquisition

We all know, for instance, that *the* is the most frequent word in English. What word is in second place? The answer is, surprisingly, corpus-specific. Frequency is typically measured using a large text corpus, but frequency calculations may differ, depending on which corpus is used to calculate frequency. Each corpus has a bias, based on the selection criterion that was used to decide which texts would be included (or excluded). This can have significant effects on the relative frequency rankings of words. For instance, the words *of, and,* and *to* are all candidates for second place, and their corpus frequencies are relatively similar. The words are ranked in that order for the Wikipedia and Project Gutenberg (literary text) corpora, but *to* moves to second place for the GigaWord corpus (newspaper text) and a corpus of essays written for the GRE assessments. The shifting of rankings becomes even greater for words that are less frequent. As the variety and size of corpora increase, the stability of a word's frequency across a range of corpora may become a significant indicator of its usefulness. The fact that a word appears in a wide variety of corpora suggests it is generally useful across a variety of contexts. Words that are consistently ranked high in frequency can be considered to be more generally useful than words that are less consistently frequent across corpora.

Somewhat different issues arise with measures of the age of acquisition. There are online systems that can inexpensively gather large amounts of information from people. This crowdsourcing technology has made it much easier to create large databases of subjective age-of-acquisition ratings, such as those produced by Kuperman et al. (2012). However, these subjective ratings are far more reliable as a relative measure of word difficulty than as actual estimates of when words were mastered by most of a target population, and it remains very difficult to collect objective measures from students for age of acquisition for a large number of word meanings. By now, the LWV is well out of date, since the demographics of the U.S. school-age population, and for that matter aspects of English vocabulary, have changed significantly since LWV estimates were collected. For example, *microwave* is defined in the LWV as a *high-frequency wave,* but it is typically understood by contemporary students as a type of oven. *Internet* and *iPod* are among the most frequently used words in modern student essays, but they do not occur in LWV. Updating the LWV remains an extremely challenging proposition, and thus far, no one has undertaken it. Moreover, except for the LWV, no other resources provide estimates of the age of acquisition for word meanings rather than for undifferentiated words.

As useful as frequency is as an indicator of word difficulty, there are some inherent problems with frequency calculations that conceptually limit its utility. The chief issue is that frequency is calculated from strings of letters and thus it does not make distinctions between words' senses or homophones. For example, the frequency for *game* reflects both its meaning of *playing a game* as well as *hunting game,* even though the latter meaning is much less frequent. More subtly but just as important for selecting words to teach or test, the frequency of a word does not tell you how often the sense of a word appears. For example, the frequency of the word *transfer* does not reflect how often that word was used to mean concrete senses, such as *transferring* groceries from the car to the kitchen or a more abstract sense of *transferring* power or authority. Frequency can also be affected by a word's appearance in multiword expressions. The frequency of *hot* includes expressions like *hot dog, hot line, hot rod,* and *in hot pursuit*.

Another issue with the reliance on frequency is that frequency cannot provide information about the role or importance of a word, either in the language as a whole or in a particular context. For example, the following pairs of words have nearly identical frequencies, according to the Corpus of Contemporary American English database (Davies, 2009): *piglet* and *metacognitive*; *entitled* and *lip*; *confirmed* and *eighteen*; and *dismissed* and *helicopters*. A cautionary note to heed when interpreting labels such as "high-frequency words" and "rare words"

is that frequency is relative, since nearly every scholar has set different thresholds for defining words as high or low frequency and there is no agreed-upon set of "high-frequency words." For example, Nation (2001) notes that the words he categorizes as low frequency include those that a reader will rarely encounter, in addition to words very close to the threshold for high-frequency words. In general, no matter how the category of low-frequency words is defined, it will contain a huge number of words because most words in the language are not used frequently. Frequencies vary across corpora and scholars vary in the thresholds they assign, so the same word may be classified as high frequency in one analysis and lower frequency in another. As a result, the categorization of words based on frequency is flexible, although the overall trends are likely to remain the same.

Alternative Approaches to Word Selection

In light of the limitations of frequency as a guide to word selection, a number of strategies have been suggested that prioritize words for instruction, based on additional kinds of selection criteria. One approach is to group together words belonging to the same morphological word family (Bauer & Nation, 1993; Nagy, Anderson, Schommer, Scott, & Stallman, 1989; Nagy & Hiebert, 2010). In particular, instruction focused on common root words has generally proven more useful than instruction focused on affixes. For example, teaching *organize* and its family members such as *disorganized, organization, reorganize* allows students to learn other forms that can be taken with this root. Teaching specific prefixes and suffixes in isolation is less effective, as not all affixes can be combined with all roots (e.g., adding *bi-* or *-ful* to *organize* does not work). A focus on word families allows a teacher to address the relationships between the family members (as we discussed with *interpret* in Chapter 2) and can help with decisions about which words to assess and teach. The larger word families provide a rich set of relationships between the variant forms and can act as exemplars for how the semantics of the morphemes combine.

Another kind of word relationship that has been used to group words is based on categories or topics, including communication; emotions and attitudes; character traits; social relationships; social categories to which people can belong; actions people can perform; comparisons/values; the body and bodily health; places/dwellings; physical attributes; descriptions of nature, of machines, and of aspects of social systems. Hiebert and Cervetti (2011) advocate this approach and call the groupings *vocabulary mega-clusters*. They suggest that teachers identify such clusters and use them to teach students how to relate words with

common meaning elements across texts. This approach is especially useful when considering that individual words will appear relatively rarely. For instance, a teacher might have students analyze multiple stories that focus on a specific cluster of emotion words, such as words related to *fear.* This might cover a broad range of words from different parts of speech, such as *caution, fright, fear, shock, terror* (nouns); *scare, frighten, petrify, wince, flinch, tremble* (verbs); or *afraid, frantic, desperate* (adjectives). Developing graphic organizers that link vocabulary into clusters, Hiebert argues, would help to develop a deep understanding of vocabulary that focuses around critical concepts in narratives.

The approach of teaching semantically related words together is not uncommon in the vocabulary literature; however, it is not without its problems. Nagy and Hiebert (2010), reviewing the literature on semantically driven approaches to vocabulary teaching, warn of a risk in teaching semantically related words together: the words may be confused in memory and therefore learned less well than if they had been taught in isolation (Tinkham, 1997). Nagy and Hiebert suggest that it may be better to teach topically related vocabulary together (i.e., words like *law/police* or *learn/school*) than to teach categorically related vocabulary together (i.e., words like *peach* and *apricot*), which may be more susceptible to confusion.

Graves et al. (2014) advocate a system for selecting words for instruction from the texts that students are reading using multiple criteria. They distinguish four categories of unfamiliar words:

1. *Accessible words*—"more common or higher-frequency words that are not likely to be understood by students who have limited vocabulary knowledge" (p. 336). These words correspond roughly to the words that Biemiller recommends for direct instruction in order to help students who need to catch up with their peers.

2. *Valuable words*—words that "have broad, general utility for students' reading and writing and thus have enduring importance" (p. 336). This group of words corresponds roughly to Beck and McKeown's Tier Two words, although what counts as a valuable word is determined in relation to both the text and the sophistication of students' vocabulary.

3. *Essential words*—words that may or may not be "valuable" in general, but are crucial for comprehending a specific text that students are reading.

4. *Imported words*—words that "help students analyze and extend what they learn in a text" (p. 336). These words capture key elements but may not actually occur in the text. Rather, they have morphological and semantic connections to the texts that students are reading and

could thus help by creating stronger connections among sets of related words.

Efforts to develop alternative word selection procedures have been chiefly motivated by a desire to help students see relationships among words and present words that are optimally useful for students' future comprehension.

Subjective Judgment

Each approach to word selection that we have reviewed ultimately relies on subjective judgment. This is particularly the case when trying to select words that will play important roles in students' literacy futures. Biemiller and Boote (2005) acknowledge this state of affairs in two ways. First, they discuss the lack of clear criteria for distinguishing teachable words and defining words that may be considered too difficult, saying that teachers are "left with testing and some uses of intuition for identifying word meanings for instruction" (p. 4). Second, they mention in their study the need to evaluate "word importance" to distinguish words that are the most useful to learn. Similarly, Hiebert (2005) acknowledges the judgment component in discussing implications for teachers like pointing out the relative utility of words that appear in the same text, for example, *checkpoint* and *cautiously* (pp. 260–261).

Using the judgments of experienced teachers to capture the words that students are likely to find difficult is time-consuming but yields a rich picture of vocabulary demands from the words that students are typically exposed to over the course of a year. Scott and her colleagues (Flinspach et al., 2012; Scott, Flinspach, & Vevea, 2011) created a word bank that reflects widely used classroom materials as source materials for determining grade-level vocabulary, based on 19 commonly adopted fourth- and fifth-grade math, science, English language arts, and social studies textbooks and 21 Newbery Award–winning novels. A team of eight experienced elementary teachers, used to working in classrooms with large percentages of English learners, combed through these materials to find words that they considered "unfamiliar" or "conceptually new" for typical students in each grade. The resulting list yielded more than 39,000 individual word forms. Winnowing this list down to words that occurred more than once, with attention to morphological patterns and polysemy, allowed them to create a teacher-tested pool of challenging vocabulary from across the curriculum. One benefit of this exercise is that it exposed the vast array of words that we expect students to be able to understand in grade-level materials and how traditional techniques may be underestimating the task. The teachers identified words

used in unfamiliar ways (*board* as in *board a train*) and as different parts of speech (*respect* as both a noun and a verb), as well as infrequent words in various subject areas (*latitude, luxurious, yucca, vaporize*). An analysis of the almost 3,000 words that occurred at least twice from the math and science textbooks revealed that 75% were beyond the 2,000 most frequent words in English—the Academic Word List only captured 13% of the words identified by the teachers. The agreement rate between teachers was 72%, but the overlap between teacher judgments and publisher-identified words was only 31%.

Selecting words for instruction is inherently a fuzzy problem space. This presents difficulties in decision making, but the positive side is that, to some extent, it does not matter which words are taught. That is, although we can identify sets of words that students will need and will likely encounter frequently, there is no perfect set of words that students must know. The set of requisite words depends on the students, their conceptual understanding, the texts they will read, and the evolving use of language.

The impossibility of identifying the perfect set of words means that teachers have and should take liberty in making selections, using a combination of informed judgment based on information about word frequency and other approaches for selection, and then develop a sense of their students' language needs. Astute word selection means attention to how words are used, which words seem to turn up repeatedly in classroom materials, and which words students seem to trip over or find difficult.

The other side of this decision making is that decisions must be made! As we have noted before, there are far too many words in the language to teach them all. There may also be far too many unfamiliar words in a text to teach them all, and attempts to do so may simply make the learning task so overwhelming that students may not learn the words they need the most. As a quick exercise in developing such word sensitivity, consider the following set of words from Harper Lee's *To Kill a Mockingbird,* as identified on *vocabulary.com*: *abide, evasion, kin, sound* (as in *sound mind*), *resentment, pauper, sweltering, trudge,* and *apprehensive*. Which are best to teach? We would suggest that a teacher might choose to prioritize some of these words over others, as follows:

Teach	**Possibly teach**	**Don't teach**
evasion	*abide*	*kin*
resentment	*sound* (as in *sound mind*)	*trudge*
apprehensive		*sweltering*
		pauper

Evasion, resentment, and *apprehensive* are good, general words that can turn up in many contexts. They are also complex words containing multiple morphemes that will support future discussions of word families and affixes. *Apprehensive* also provides the potential for a discussion of multiple word senses, especially physical and abstract mental senses, with one sense of its root, *apprehend,* meaning literally to capture someone. The words in the "Teach" column are prioritized because they are useful for multiple reasons, such as both exhibiting morphological complexity and secondary senses.

The words in the "Don't teach" column (*kin, trudge, sweltering,* and *pauper*) are rarer and more specialized than the words in the "Teach" column, and also less morphologically productive, which means that they are less likely to be useful when students encounter a new text. They are the kind of words that may be briefly explained during class discussions, and which students should be encouraged to look up on their own, but are probably not worth sustained class time. Not teaching them, however, does not mean that they are useless or that students will never encounter them again; they are more concrete in meaning and thus students are much more likely to learn these words on their own. The "Possibly teach" words fall somewhere in the middle—they are not concrete, but not as morphologically rich. Both words have multiple senses, so this might be a reason to include them; yet they seem simpler in meaning than the recommended group.

Another way to treat words like *kin, trudge, sweltering, pauper, sound,* and *abide* is to recognize them as representing concepts that students will mostly know and discuss them in ways that help students recognize relationships to other words. Having students collect words in their own reading related to known concepts such as *hot,* ways of *walking,* and terms that describe *wealth* and *poverty,* provides a set of words from books that students are interested in knowing and using. Teaching then becomes less about selecting particular words and more about recognizing relationships among words and building on what students already know. Words such as *sound* and *abide* can engender a discussion about polysemy, drawing on information about nouns and verbs and how one might be able to infer which sense is being used in a passage.

Multiword Expressions

In addition to individual words, multiword expressions (MWEs) also should be considered in decisions about vocabulary instruction. These expressions vary along a continuum from idiomatic expressions to phrases composed of two or more words that have entirely predictable meanings. The extent to which an expression is idiomatic determines

whether it must be learned individually, as a whole phrase, or whether it can be inferred from the words that it comprises (Ellis, 2003). Academic English, in particular, requires learners to acquire a large inventory of idioms, collocations, and lexical bundles (Coxhead & Byrd, 2007). For English learners, mastering academic vocabulary is part of the transition from intermediate to higher levels of academic English proficiency (Eskildsen, 2008; Yorio, 1989).

As mentioned in Chapter 2, we know very little about how MWEs are learned developmentally. We do not have an exhaustive inventory, since standard dictionaries only contain a fraction of the MWEs in common use and are inconsistent about which expressions are included. There are no resources that a teacher can use to provide guidance about what should be taught and assessed. Nevertheless, such expressions make up at least half of our vocabulary—they are a source of student errors, and they are crucial for fluency. Some factors to consider when deciding which MWEs to teach might be their relationship to a particular discipline or to academic learning, their dispersion or frequency in English, and whether they can be used to point out particular patterns of interest.

Words with a Latinate origin are associated with academic vocabulary and with a higher degree of linguistic competence (see Corson's [1985] *The Lexical Bar*). For example, *vice versa, status quo, bona fide, carpe diem, ad hoc, ipso facto, Homo sapiens, prima facie, habeas corpus, cum laude, pro bono,* and *modus operandi* are among the most frequent expressions found in a variety of corpora. Many of these words are highly associated with the legal profession and many of them are a part of Tier Three vocabulary. In studying a particular subject, such as social studies, it might be worth drawing students' attention to these MWEs and their origins.

MWEs often use restricted secondary senses of commonly known words, creating an issue for students who know only the primary meanings of words. For example, *real estate* does not use *real* in the sense of *reality* or *real number.* Because students are often unaware that words can be used with secondary senses, MWEs can be confusing, as they exhibit a variety of senses. For example, *Milky Way* has two completely different senses, as a candy bar and as a galaxy; *garbage collector* has both the literal meaning of collecting garbage and a figurative meaning in computer science as a form of memory management; and some MWEs can have systematic secondary senses such as *New York,* which can refer to both a location and an entity (as in "New York passed a law").

MWEs can also have idiomatic meanings that cannot be understood by understanding their component words. We can understand *gravity of the offense* if we understand the secondary sense of the word *gravity* as extreme seriousness. However, this differs from expressions like *across*

the board, clear the air, and *red herring*, where an understanding of the component words does not allow the reader to understand the meaning of the expression. The *red* in *red herring* is not a secondary sense, just as we do not think of *hot* in *hot dog* as a secondary sense.

Identifying patterns associated with MWEs would promote a metalinguistic understanding of how English works. One such pattern is that phrasal verbs and idioms can be broken up into nonadjacent parts in a sentence. For example, the phrasal verb *look up* can occur as *We looked up the word* or *We looked the word up*. Similarly, idioms can be broken up into nonadjacent parts in a sentence, such as *calling someone to account* or *letting something slide*. Like variations within morphology, MWEs can be organized into families of patterns, such as the subset of phrasal verbs that expresses the concept of completion (e.g., *eat up, drink up, grow up*).

Morphology is useful in understanding restricted contexts of meaning for MWEs. The words *preoccupy* and *preoccupation* are highly related because they refer not to occupying in general, but having one's time occupied with an activity or interest. We do not talk about preoccupying a seat in the theater or say that an army sent in advance forces to set up a preoccupation. In other words, a prefix can be applied to a single sense of the base word, just as an MWE can create a restricted context in which only one secondary sense of a component word is relevant (Krovetz, 1993).

Earlier in the chapter, we divided individual words into three categories: "Teach," "Possibly teach," and "Don't teach." A similar division can illustrate some of the principles for selecting specific MWEs for attention.

Teach	**Possibly teach**	**Don't teach**
vice versa	*miles per hour*	*good morning*
ad hoc	*life expectancy*	*blue sweater*
bona fide		*car door*
eat up		
soup up		
carbon monoxide		
Civil War		
standard deviation		
scuba diving		
Herculean effort		

The first three expressions, *vice versa, ad hoc,* and *bona fide,* were selected because they have Latinate origins and because they have high

dispersion. The next two, *eat up* and *soup up,* were chosen to illustrate a pattern and exceptions to that pattern. With many verbs, the particle *up* implies that the action described by the verb has been completely accomplished (*drink up/drink, finish up/finish, grow up/grow,* and so forth). There are exceptions to the pattern, which do not have the same relationships with the verb, such as *soup up, beef up, gear up*. It is useful both to help students recognize general patterns, but also recognize that there are limitations and exceptions where the patterns do not apply.

Words in subject areas, such as science or social studies, offer opportunities to talk about MWEs, for example, expressions such as *carbon monoxide* or *Civil War. Carbon monoxide* is proposed because of the morphemes *mono* and *oxi*. Such decomposition is especially important to decoding the vocabulary of science and technology. *Civil war* is proposed because the Civil War was not *civil* in the sense of a civil discussion. *Standard deviation,* a term in statistics, illustrates the same point; there is a different sense associated with a component word.

We propose the next item on the "Teach" list, *scuba diving,* because it provides an opportunity to describe the concept of acronyms and how acronyms can evolve to become words themselves. The word *scuba* is an acronym for *self-contained underwater breathing apparatus*. The same is true for *radar* (radio detection and ranging) and *laser* (light amplification by stimulated emission of radiation). *Herculean effort* is suggested because of a connection with Greek myths, which is an often-recommended aspect of literature study. It is a particularly useful choice because it has a family of terms associated with it (*Herculean task, Herculean accomplishment,* and *Herculean feat*) and because it provides an opportunity to mention other expressions in which names are included (e.g., *Socratic dialogue, Platonic ideal,* and *Euclidean geometry*). We exclude expressions that can be easily understood by understanding the component words.

Beyond Words: Attending to Contexts and Word Patterns in Instruction

A possible consequence of thinking in terms of which specific words to select for instruction is that it may reinforce the idea that a person's vocabulary is a list. A major focus in this book, however, has been to offer a different conception of vocabulary, based on patterns of language use. This conception, taken from modern linguistic theory, implies that we should attend not only to selecting important words to teach, but also to the kinds of contexts that surround words in vocabulary instruction (Goldberg, 1995a, 1995b, 2006, 2016; Tomasello & Tomasello, 2009).

In usage-based theories, learning is driven by the words that serve as prototypical cases around which language learning can coalesce.

Consider the word *asylum,* which is representative of one of the major patterns we have discussed and exhibits multiple senses (polysemy). The sentence *The refugees sought asylum in a foreign country* allows a learner to generalize to the meaning that *asylum* has in typical social studies contexts. The example places the word in a typical syntactic construction (*seek asylum*), which implies that *asylum* is something that one looks for or obtains; indicates who typically seeks *asylum* (*refugees*); and indicates the typical location in which *asylum* may be obtained (a foreign country). The syntactic pattern of the context also suggests what synonyms of *asylum* might be—*seek refuge, seek safety.*

Including generative patterns as part of instruction essentially switches the focus of instruction from words that play an important role in the language to linguistic patterns that underlie effective vocabulary knowledge, but that may or may not include specifically targeted words. Including generative patterns in instruction does not necessarily imply direct, explicit instruction of patterns. Usage-based approaches suggest that effective vocabulary development can be encouraged by exposing students to informative contexts and enabling them to engage with those contexts in meaningful ways, for example, by talking about what a context means and what they notice about how the words fit together within the context. Noticing contexts and how they work encourages students to internalize the linguistic patterns that such contexts exemplify.

SELECTING WORDS, WORD PATTERNS, AND CONTEXTS FOR ASSESSMENT

Moving now from selections for instruction to selections for assessment, we acknowledge that one of the fundamental issues in assessment is the problem of generalization—given the sample of words tested, what conclusions can be justified about vocabulary knowledge that is broader than simply knowledge of those particular words? In Chapter 3, one of our main critiques of traditional vocabulary assessment was that the word selection principles and item types effectively limited the set of justifiable inferences about relative vocabulary size. If we want to support a broader set of inferences, we need to select words, word meanings, and contexts that will provide the relevant evidence. The question, then, is what other kinds of information do we want to capture about students' knowledge?

In this section, we first consider a perspective on selecting words that might yield broader information. We then suggest three additional

aspects of vocabulary knowledge to consider for assessment design. Two of these are drawn from our conception of generative patterns, morphological knowledge, and polysemy, and the third is a category we have discussed as a key part of the structure of our vocabulary: MWEs.

Words can be selected for assessment either in a passage context (if we obtain specific supporting texts that are appropriate for a population) or, more generally, outside any specific passage context. If we select words in a passage context, it is important to consider the purpose for which a word is being assessed, which, in turn, is related to why a teacher might choose to teach it. The categories developed by Graves et al. (2014) for selecting words for instruction might thus be usefully adapted to consider selections for assessment. For example, words that may be important for assessing students with limited vocabulary knowledge, such as English language learners, correspond to Graves et al.'s *accessible* words.

Other words might be selected because they are challenging for most students and important for reading a variety of texts in the future, which would correspond to *valuable* words. Other words may provide important information about students' reading comprehension because students may misunderstand a text solely because they have failed to understand an *essential* word. Moreover, there may be words that do not appear in the text, but which relate to the text, and which might be useful to assess to evaluate students' knowledge of the topic or theme more broadly; such words would correspond to *imported* words. These categories of words comprise four distinct reasons to prioritize words for assessment. If we do not test *accessible* words, we may miss a significant source of difficulty for struggling students. If we do not test *valuable* words, we may not know how well students' performance will generalize across texts. If we do not test *essential* words, we may not know how well students will be able to deal with the key content the text presents. In addition, if we do not test *imported* words, we cannot determine how well students will perform when they are asked to discuss or analyze text content. Depending on the purpose of a vocabulary assessment, we might choose to assess different mixes of *accessible, essential, valuable,* and *imported* words for inclusion in a vocabulary assessment.

Note that these distinctions are best treated in relation to targeted texts. If we select texts that exemplify the reading demands that we expect students in a particular population to be able to handle, it should be possible to develop effective methods to identify words from each of these categories. During assessment design, an analysis of a corpus of such exemplary texts can be used to develop lists of targeted words that could be used to support stand-alone assessments that do not make direct use of passage reading. This was the approach taken by Scott

and her colleagues in developing the VASE assessment (Flinspach et al., 2012; Scott et al., 2011). However, for many purposes, it may be most appropriate to select rich contexts and base the words to include in an assessment of those words' presence in those texts. It seems especially important to select contexts that can be used to determine whether students have generalized from instruction. It is crucial to assess vocabulary that has not been explicitly taught. In the discussion that follows, we suggest three additional aspects of vocabulary knowledge that we believe should be considered during word selection for assessment design drawn from our conception of generative language patterns. Each has implications for the choice of words, word meanings, and contexts for assessment.

Tacit Morphological Knowledge

As discussed in Chapter 2, a key source of vocabulary growth between third grade and the end of high school is the ability to combine known words and morphemes to acquire new, morphologically complex words. Success with this learning depends on morphological awareness, the ability to recognize how a new word is related to words that have already been learned. Thus, knowing the word *bribe* and the meanings of the affixes *un-* and *-able* would allow a student to infer that the word *unbribable* means someone who cannot be bribed (Anglin, 1970).

Because strong morphological awareness predicts strong vocabulary knowledge and reading comprehension, it is important to include words with various morphological patterns in a vocabulary assessment. This will allow us to assess not only tacit morphological knowledge, but to obtain information about students' conscious knowledge of morphological patterns. The type of word forms that we choose to test will allow us to gather information on the level of students' implicit awareness of morphological patterns. If we test a rare word like *sleeplessness* and discover that students know it is a noun, we have evidence they understand the implications of the *-ness* suffix. If they understand what the word means, we may have evidence that they understand not only the base word *sleep* but also the semantic implications of *-less* and *-ness*. Since the base word and the prefix are part of the core Anglo-Saxon vocabulary and since the suffixes represent highly productive syntactic patterns, a failure to understand them would represent a far lower level of vocabulary development than a failure to understand a word like *irreversibility,* which draws on opaque Latinate morphology. Latinate morphology is more complex than Germanic, and it is typically acquired later.

If learners have a grasp of tacit morphological patterns, they will be well equipped to make sense of unfamiliar words that have transparent,

productive morphological patterns. Even very rare words like *irreversibility* are accessible if a learner first knows the word *reverse,* understands the tricky nature of the *in-/ir-* prefix, how *-ble* can be added to a verb to create an adjective, and how that can then change to *-bility* to indicate a noun form. However, the successful use of morphology relies on already knowing the root—either a base word such as *reverse* or a bound root within a word such as *-nov* in *innovative.* Without that key knowledge, morphology is not an effective tool. If we do not obtain a clear sense about the range of morphological patterns with which students are able to deal with effectively, we are unlikely to have a clear sense of how well prepared they are to address the ever-increasing variety of derived words they will encounter in more advanced texts.

We have been talking about knowledge of morphological patterns as tacit—learners have the knowledge and can apply it to understand language, even though they may not be able to explain either the patterns or what they know about them. However, there is strong evidence that learners who are consciously aware of such patterns and can manipulate them are more effective at reading and word learning (Anglin et al., 1993; Carlisle, 2000; Carlisle & Feldman, 1995; Kirby et al., 2012).

Polysemy

We have introduced polysemy, multiple but related meanings of a word, as a generative mechanism that underlies much of the expansion in vocabulary knowledge that happens during the school years. Polysemous words, therefore, are prime candidates for inclusion on assessments, as the ability to understand polysemous senses and extend their meanings to novel contexts is key to successful reading comprehension. When selecting words for assessment, we cannot assume that students will know one meaning (especially secondary, less common meanings) even if they know other meanings of the word. Thus selecting words for assessment is really a matter of selecting word senses, and secondary meanings should not be ignored.

Students with strong word-learning skills may have the semantic flexibility to infer what a word means in context. For example, the sentence *She felt a lot of pressure to join the group* favors a different meaning than the dominant sense of the word, as in *He put pressure on the wound to stop the bleeding.* To understand the first sentence, people must recognize the abstract, metaphorical meaning of *pressure* as social pressure and infer that is the intended meaning. But we cannot assume that all students will be able to make this kind of inference. Many students may not be aware that this second meaning exists or they may fail to retrieve it in context.

There is significant evidence that primary senses are more easily accessed than secondary senses; that individuals can vary in semantic flexibility; and that students, at least in the primary grades, tend to recognize secondary senses of words less accurately, even when the words appear in compatible, supportive contexts. There is also evidence that the dominant senses of word meanings come to mind readily, as the default, whereas if a context uses a secondary sense, understanding it requires more effort to integrate the meaning with the context (Foraker & Murphy, 2012; Giora, 2012; Graves, 1980; Titone & Salisbury, 2004). For example, third and fourth graders are much more likely to ignore the context of the sentence *The workers are going to strike tomorrow,* relying on their knowledge of the primary meaning to choose *hit* as the meaning of the word *strike* (Mason, Kniseley, & Kendall, 1979). Mason et al. conclude that when words were assigned their primary (most frequent, prototypical) sense in context, students remembered the words accurately and assigned them the correct meaning, but that when words were assigned secondary senses, they frequently misremembered the context and assigned the word to its primary rather than its secondary sense. This tendency was significantly correlated with reading ability, such that higher-ability students performed more accurately on secondary senses, without any significant difference across grade levels.

Sullivan (2006) had a similar finding about the relationship between ability and word senses. She found that higher-ability students were able to provide explanations of relationships between senses compared with less proficient students, regardless of the grade. Her work also demonstrates that the acquisition of secondary senses accounts for much of the growth in vocabulary knowledge between third and 12th grades. This body of research makes strong arguments in favor of selecting secondary as well as primary senses of words for both instruction and vocabulary assessment and in selecting contexts that provide evidence to support word-sense disambiguation, since the context in which students read a word may provide stronger or weaker support for the inferences they need to make. If we do not design assessments to gauge semantic flexibility, we may miss critical differences among stronger and weaker students.

Multiword Expressions

Choosing to assess MWEs can help teachers determine whether the students can recognize systematic patterns, such as the notion of completion that is a part of *eat up, finish up,* and *grow up,* and can provide information about whether students have the correct understanding of

the expressions. For instance, we can ask students to paraphrase *soup up* as *modify and improve* and *beef up* as *strengthen*. We can also ask them to create "idiom books" by creating illustrations that show both the literal and the actual meaning of common idioms. Some good candidates would be *kick the bucket, red herring, raining cats and dogs, face the music,* and *hold your horses*.

FINAL THOUGHTS

So let us return to the core question of this chapter: how many words should be selected and how should those words be taught? It should now be clear that the criteria for selecting words are multiple and complex. A one-size-fits-all approach to word selection may not best serve the learners. In the classroom, much depends on a teacher's judgment and where students are in their mastery of English. However, we have identified several principles that can help teachers make effective choices.

1. Focus explicit instruction on useful, general-purpose words that will help students acquire academic English. Resources like Biemiller's Words Worth Teaching (Biemiller, 2009), Beck et al.'s (2002) Tier Two words, the Coxhead (2000) Academic Word List, and Hiebert's (2011) Word Zones may prove to be useful for this purpose.

2. Emphasize words at or slightly above grade level, which will usually be understood by the stronger students in a class. These are the critical words weaker students need to learn.

3. Pay attention to words in context in the texts students are reading. Look for teachable moments when students encounter unfamiliar or difficult vocabulary—and be alert for challenging features, such as polysemy (multiple meanings) and MWEs. Encourage students to actively identify words they need to know in the texts they are reading and teach them how to make effective use of language resources (dictionaries, thesauruses) as tools, not crutches.

4. Encourage students to think about and analyze words as a normal practice as they read, and encourage them to recognize patterns in how words are used. This includes explicitly analyzing morphology, focusing on common root words. It includes identifying lists of related words that describe the same categories and topics. It includes paying attention to the reasons for choosing one word over another and identifying important or repeated words that can help them better understand the texts they are reading.

These suggestions provide a framework for selecting words, word patterns, and contexts for instruction and assessment that can help teachers and students to navigate the sea of words. In the end, the teacher plays a critical role. Teaching vocabulary is not merely teaching lists of words, but also choosing words that extend students' understanding for multiple contexts so they can take charge and become able to expand their own vocabulary.

Guilford Publications
370 Seventh Avenue, Suite 1200
New York, NY 10001
212-431-9800
800-365-7006
www.guilford.com