What Differences in Narrative and Informational Texts Mean for the Learning and Instruction of Vocabulary

Elfrieda H. Hiebert
TextProject, Inc.

Gina N. Cervetti
University of Colorado at Boulder

TextProject

textproject.org

What Differences in Narrative and Informational Texts Mean for the Learning and Instruction of Vocabulary

Elfrieda H. Hiebert
TextProject, Inc.

Gina N. Cervetti
University of Colorado at Boulder

Abstract

This project examined the words selected for instruction from fourth-grade English/Language Arts (ELA) and science programs with the goal of describing the unique words in these two text types. Seven features of the words were established: (a) length, (b) frequency, (c) frequency of a word's morphological family, (d) familiarity, (e) dispersion (i.e., how frequently a word appears across subject areas), (f) conceptual complexity, and (g) semantic relatedness. Analyses showed differences on all features except for the frequency of morphological families and dispersion. Narrative vocabulary was more familiar but less frequent than science vocabulary, but science words were longer, more conceptually complex, and more semantically related than narrative words. These differences lend themselves to different instructional approaches. In science, where unique words are conceptually complex, students benefit from extensive discussion and demonstrations. Because the unique words of narrative texts represent fairly familiar concepts, instruction should emphasize the ways in which authors vary their language.

Contents

What Differences in Narrative and Informational Texts Mean for the Learning and Instruction of Vocabulary

WE BEGIN WITH FOUR STATEMENTS ABOUT INFLUENCES ON VOCABULARY instruction in schools. First, vocabulary is central to the comprehension of texts (Davis, 1942; Thorndike, 1973). Second, the vocabularies of students when they enter school vary substantially (Hart & Risley, 1995). Third, the number of words in English is huge (Leech, Rayson, & Wilson, 2001). And fourth, the amount of time in schools is limited (Fisher et al., 1980). All of these features combine to create a challenging situation for educators who aim to select vocabulary strategically in order to lessen the gap between the haves and the have-nots (Nagy & Hiebert, 2010).

Unfortunately, it appears that the choices made in schools regarding vocabulary are often not strategic. In elementary schools, large blocks of time are devoted to reading/language arts instruction where, despite claims of increased amounts of informational texts within core reading/language arts programs, a narrative stance has continued to direct the selection of vocabulary and the form of vocabulary instruction (Norris, Phillips, Smith, Baker, & Weber, 2008). Whether the text is an informational or narrative one, teachers' guides of core reading programs recommend instruction of a handful of words for each text. Typically, these words are treated in a similar manner—each is defined, discussed, and read in the context of a sentence from the text. Usually, the words are unrelated to one another but have been picked because of their perceived importance to the content of the text. For example, words that describe the feelings of a group of storm chasers watching an approaching hurricane (e.g., *anxiously, scarier, worried*) might be recommended as focus words rather than words having to do with weather forecasting (e.g., *anemometer*) or storm conditions (e.g., *storm surge*).

Such a perspective fails to recognize the differences in the vocabularies of narrative and informational texts. Typically, the registers of oral and written language are recognized as unique, but these differences pale relative to differences in the features of narrative and informational genres. Through multidimensional analyses of spoken and written language samples, Biber (1988) concluded that particular types of speech and writing can be quite similar. For

example, an oral presentation at a meeting of a scientific society will vary considerably from a conversation between two friends over dinner. The vocabulary of a novel that includes substantial amounts of dialogue may have more in common with the dinner conversation than with the scientific report.

In this chapter, we examine the differences between the target vocabularies of an English/Language Arts (ELA) program that is dominated by narrative texts and a science program consisting of informational texts. Our goal in this chapter is to accomplish three purposes: (a) review what is known about the differences in the vocabularies of unique words in narrative and informational texts, (b) verify these differences in an analysis of the words from an ELA and a science program, and (c) present suggestions as to what uniquenesses in the vocabularies of different text types mean for instruction.

What Is Known About the Differences in the Vocabularies of Narrative and Informational Texts?

To understand differences in the vocabularies of different subject areas requires a foundation in the features of words in written English. Differences in words have been identified on numerous dimensions, including but not limited to their length, part of speech, and etymology. To describe the differences of the topic-specific words in different genres, we focus on three criteria: (a) frequency of the word and its morphological family, (b) conceptual complexity and familiarity, and (c) relatedness within a thematic or semantic network of words.

Frequency of Words and Their Morphological Families

The approximately 750,000 words in the British National Corpus (Leech et al., 2001) can be sorted into three groups on the basis of frequency: (a) highly frequent, (b) moderately frequent, and (c) rare. The first group is made up of approximately 1,000 words that account typically for two-thirds of the total words in a text. The first row in Table 1 shows the high-frequency words within 50-word excerpts from two fourth-grade texts, one a narrative text (Gerson, 1994) used in Afflerbach et al. (2007) and the other an informational text (Cooney et al., 2006). Words such as *object, energy,* and *matter* in the second column of the first row of Table 1 show that all of the 1,000 most frequent words are not simply glue words such as prepositions, pronouns, and question words. Some of the words in this group are there because they have multiple meanings. In science, words such as *energy* and *matter* take on quite precise meanings that differ from their common use. With only the words from the 1,000 most frequent group (as is the case in Row 1, Table 1), a reader gets the sense that the text is about objects, energy, heat, and movement but does not have sufficient context to know precisely how these terms fit together.

A group of approximately 4,750 words appears with moderate frequency in written language—10 to 99 times per million words. Examples of words within this group are given in the second row of Table 1. Although specific concepts

TABLE 1

Distributions of Words by Frequency in Exemplar Narrative and Informational Texts

	Narrative Text	Informational Text
High	Iemanja's daughter loved **her** husband, and **she** loved the magic **of** daylight that he showed her; the shimmering **sand** of the beach, **the** rows **and** rows of cocoa **and** sugarcane baking **in** sunlight, **and the** sparkling jewels **and** feathered costumes worn **in** harvest festivals	Particles in an object move because they have energy. As an object becomes hotter, its particles move faster. As the object cools, the particles move more slowly. Thermal energy is energy due to moving particles that make up matter. We feel the flow of thermal energy as heat.
Moderate	Iemanja's **daughter loved** her **husband,** and she **loved** the **magic** of **daylight** that he showed her; the shimmering sand of the **beach,** the **rows** and **rows** of cocoa and sugarcane baking in **sunlight,** and the sparkling jewels and **feathered** costumes worn in **harvest** festivals.	**Particles** in an object move because they have energy. As an object becomes hotter, its **particles** move **faster.** As the object cools, the **particles** move more slowly. Thermal energy is energy due to moving particles that make up matter. We feel the **flow** of thermal energy as heat.
Rare	**Iemanja's** daughter loved her husband, and she loved the magic of daylight that he showed her; the **shimmering** sand of the beach, the rows and rows of **cocoa** and sugarcane baking in sunlight, and the sparkling jewels and feathered **costumes** worn in harvest **festivals.**	Particles in an object move because they have energy. As an object becomes **hotter,** its particles move faster. As the object **cools,** the particles move more slowly. **Thermal** energy is energy due to moving particles that make up matter. We feel the flow of **thermal** energy as heat.
	Iemanja's daughter loved her husband, and she loved the magic of daylight that he showed her; the shimmering sand of the beach, the rows and rows of cocoa and sugarcane baking in sunlight, and the sparkling jewels and feathered costumes worn in harvest festivals.	Particles in an object move because they have energy. As an object becomes hotter, its particles move faster. As the object cools, the particles move more slowly. Thermal energy is energy due to moving particles that make up matter. We feel the flow of thermal energy as heat.

are present (e.g., *Africa, France, Mexico*), the majority of words in this group represent common concepts (e.g., *lakes, villages, desert*). At times, words that represent common concepts (e.g., *flow*) can take on specific meanings, as is the case in the science text. With the addition of this group of moderately frequent words, readers can gain the gist of the text, such as the daughter's love of the light in the narrative example. Sufficient context is available to understand that a common word such as *flow* takes on a specific meaning in the science text.

Beyond the 1,000 highly frequent words and the approximately 4,500 moderately frequent words, the remaining words in written English—up to 745,000 words according to the British National Corpus (Leech et al., 2001)—appear less frequently. As can be seen in the narrative excerpt in the third row of Table 1, some of these words are names of people. Others are representations of known concepts that authors use to give nuance to their writing—*shimmering, sparkling.* Still others are concepts such as *thermal* that are unique to domains. Approximately 15,000 of these words appear from 1 to 9 times per million.

The remaining words of English—approximately 97% of the words in the language—can be expected to appear less than once per million words of text.

Many words in this group of approximately 725,000 rare words are archaic (e.g., *bap, snell*). The *Oxford English Dictionary* (Simpson & Weiner, 2009) identifies approximately 425,000 active words in English. When words are considered as morphological families, rather than as individual words, the volume of words is approximately five to six times smaller (Nagy & Anderson, 1984). Viewing the frequency of a word as a function of the size of its morphological family is justifiable in that nouns and their plurals, as well as conjugations of verbs, share a representation in the mental lexicon (Sereno & Jongman, 1997; Stanners, Neiser, Hernon, & Hall, 1979). Developing and struggling readers can be challenged by multisyllabic words, which most morphologically derived words are (Nagy, Berninger, & Abbott, 2006). Word meanings, however, prove the greatest challenge to students' comprehension—even more than features such as length and frequency (Nagy, Anderson, & Herman, 1987). A word such as *energy* has a specialized meaning in a physics text (e.g., $E = mc^2$) that differs from the meaning communicated in daily conversations (e.g., *I don't have the energy to cook tonight*).

Conceptual Complexity and Familiarity

The essence of language is its meaningfulness, and it is the word that represents unique entities. Particular words may appear infrequently in written language, but they may be quickly recognized in a text because they are highly concrete (e.g., *skateboard, mirror*) or can be easily understood from contextual use. An instance of the latter is illustrated by the use of the word *madragada* in the following sentence from Gerson (1994): "In Brazil the early morning is called the madragada."

Jenkins and Dixon (1983) identified four relationships between a learner and a new word: (a) unknown word but a known concept that can be expressed succinctly (altercation/argument); (b) unknown word with a simple synonym but student does not know the concept referred to by the synonym (arcane/obscure); (c) unknown word that does not have a simple synonym but can be described through experience (e.g., odometer/thing on speedometer that tells how many miles you have gone); and (d) unknown word that does not have a simple synonym and for which students do not have extensive experiences (e.g., *legislature*). The density with which unknown words of the fourth type appear in texts is likely a strong influence on students' comprehension (Sternberg & Powell, 1983). Students may be able to establish the meaning of a conceptually complex word with an unknown meaning in a paragraph. Their comprehension may be compromised, however, when the ratio of unknown to known words reaches a particular threshold. They may also be unable to deepen their knowledge of new words when texts are dense with unknown words.

A study conducted by Nagy et al. (1987) confirms the hypothesis that conceptual complexity of words influences students' ability to understand unknown

words while reading. Third, fifth, and seventh graders were given texts that had unknown words that varied in conceptual complexity. Using a scheme for conceptual difficulty similar to that proposed by Jenkins and Dixon (1983), Nagy et al. found that conceptual difficulty was the only word feature from among several (including length, part of speech, and morphological complexity) that was significantly related to students' ability to understand the word's meaning in context. The properties of texts that most influenced students' learning words from context were the proportion of unfamiliar words that are conceptually challenging and the average length of unfamiliar words (an indicator of morphological complexity).

Semantic Relatedness

Words enter the lexicon as humans make distinctions about features of the world around them, both internal and external. Consider, for example, two words that have been officially recognized by lexicographers over the last year (Oxford Dictionaries, 2010): *neuroprotective* and *spyware*. Words such as these are not the product of random word generators but of human beings making unique distinctions among entities or experiences in their environments. Words are parts of a richly interconnected network (Entwisle, 1966; Levelt, Roelofs, & Meyer, 1999). Common relationships among words include semantic classes (e.g., *eggs/food*), collocation of words that commonly occur together (e.g., *a dozen eggs*), superordination (e.g., *sedimentary/rock*), and synonyms (e.g., *glittering/sparkling*). Moss, Ostrin, Tyler, and Marslen-Wilson (1995) describe additional ways in which words are related within the mental lexicon, such as part-whole (*branch/tree*), instrumental (*broom/floor*), and scriptal (*hospital/nurse*).

Within a curriculum area such as science (Marzano, 2004), words are clustered within thematic groups. For example, within the vocabulary recommended for science instruction in standards documents for Grades K-2 are words and phrases associated with weather (e.g., *weather conditions, weather patterns, seasonal change, precipitation*). On the list of words that Marzano (2004) identified as the vocabulary in ELA standards documents were the words that are typically used in instructional conversations led by teachers—words such as *vowels* and *consonants*. Such words are not the ones that are found in the ELA texts read by students, unless the texts are workbooks. In a science curriculum, vocabulary identified within standards documents would be expected to appear in texts and lessons. It would be unusual, however, within ELA for a story to be about *vowels* or *consonants*.

Words that appear among the moderately frequent and rare words of the narrative text in Table 1 (e.g., *feathered, loved*) do not appear within standards documents as recommended concepts. The typical response to this observation is that the variety in the words used in stories is so substantial that systematic selection of vocabulary in ELA standards documents is impossible. However, if literary words such as *costumes, shimmering, festivals,* and *feathered* are seen

as members of larger semantic clusterings of ideas, a systematic and cohort approach to the selection of words may be possible, if not the identification of specific sets of words.

A proposal based on research about semantic connections suggests a way in which vocabulary might be taught. This proposal came from Marzano and Marzano (1988) who organized 7,300 words from word lists for elementary students into 61 superclusters of words (e.g., types of motion) that were further broken into 430 clusters where words had closer semantic ties (e.g., *taking/ bringing* and *tossing* within the motion supercluster). The clusters were made up of 1,500 miniclusters such as the eight within the *taking/bringing* mini-cluster (*take, return, get, send, remove, put, deliver, import*). Such a system has support in the research literature where teaching groups of words that are semantically related—such as *law/police, leaf/tree,* and *learn/school*—has proven to impact learning positively (Tinkham, 1997). Nagy and Hiebert (2010) suggested that similar words might be taught gradually with a known member of a semantic set serving as an anchor because teaching words that are too similar in meaning also can interfere with student learning (Tinkham, 1993; Waring, 1997). In other words, all of the words in one of the Marzano and Marzano (1988) miniclusters would not be taught simultaneously. Words in texts that share semantic clusters and miniclusters would be taught in relation to known words within the clusters and miniclusters. For example, *shimmering* and *sparkling* might be taught in relation to the likely known word *shining*. Nagy and Hiebert emphasized that the goal of a curriculum is to teach concepts, not just individual words, and that concepts have relationships to one another.

Research on Differences in the Vocabularies of Narrative and Informational Texts

The words from the exemplars in Table 1 illustrate how words of moderate and rare frequency which are unique to either informational or narrative texts (i.e., appear in only one of the text types) represent different types of concepts. Armbruster and Nagy (1992) identified three differences between the unknown words of narrative and informational texts: (a) knowing these words is likely more crucial to getting the gist of informational texts than of narrative texts; (b) these words are likely more conceptually challenging in informational texts than in narrative texts; and (c) the words in informational texts are likely more interrelated thematically than those in narrative texts. However, empirical verification of these differences has been limited.

Although the presence of different types of vocabulary has been identified as one of the features that distinguish genres from one another (Biber, 1988), descriptions of the features of vocabulary in narrative and informational texts used in elementary schools have been insufficient. We have found only a single study that has analyzed differences between the words in narrative and content area texts. This study—by Gardner (2004)—was focused on the number of nonfrequent words that were shared or unique to narrative or informational

texts drawn from the same three themes (mummies, mystery, and westward movement). After Gardner had eliminated the words on the General Service List (GSL; West, 1953), or the University Word List (Coxhead, 2000), there were 23,857 unique words (from a total sample of approximately 1.4 million words). Of these 23,857 words, 42% appeared only in narrative texts and 30% appeared only in informational texts. The remaining 6,566 unique words were analyzed to determine how many appeared 10 times or more within both genres, a level that Gardner identified as a sufficient number of repetitions for meaningful acquisition. This group of shared unique words with 10 or more repetitions was 233. What is clear from this analysis is that the vocabularies that appear in these different genres have limited overlap, even when the texts have been chosen to represent the same topics. Gardner (2004) did not conduct additional analyses to determine what distinguished the three groups of unique words. Without greater understanding of the characteristics of the many words that are unique to one or the other genre, publishers and educators are left uncertain as to how words should be chosen differentially and what these features mean for instruction. To ameliorate this gap, we conducted an analysis of the features of words identified for instruction in ELA and science programs.

What Differences Were Apparent in an Analysis of the Vocabularies of Narrative and Informational Texts?

Although scholars conclude that the vocabularies of narrative and informational texts have unique characteristics (e.g., Armbruster & Nagy, 1992), descriptions of these differences are limited. Consequently, we conducted an analysis of the features of the vocabularies of these two types of texts for this chapter. We analyzed the features of all of the words that have been identified for instruction and assessment within both an ELA and a science program. We also analyzed the words from exemplar texts from each program.

An Analysis of the Word Features

Our analysis of the word features of narrative and informational texts focused on all of the words that are designated for instruction (and subsequently assessment) from the fourth-grade ELA (Afflerbach et al., 2007) and science (Cooney et al., 2006) programs of the same publisher (Scott Foresman) for the entire school year. The ELA program had 209 words, and the science program had 207.

A prefatory comment is needed about the attribution of narrative to the vocabulary and texts of the ELA program. As has been documented recently (Norris et al., 2008), the genres evident in current core reading programs include informational texts focusing on science and also social studies. Although potential exists for developing the vocabulary of content areas with these texts, Norris et al. reported that the recommended instruction and assessment is more appropriate for literary texts than for informational texts. Our perusal of the

vocabulary with the ELA program confirmed the findings of Norris et al. For example, in a text on the tracking of hurricanes, vocabulary that mirrored the vocabulary in narratives (e.g., *expected, shatter, destruction*) was highlighted rather than the scientific vocabulary present in the selection (e.g., *anemometer, meteorologists, tornadoes, satellite, storm surge*). Although a significant portion of the texts in the ELA program came from content-area sources, criteria for selecting vocabulary from these texts appeared to be the same ones as those used for narrative texts.

Although the number of lexical items identified for instruction was similar across the ELA and science programs (209 for the former; 207 for the latter), there was a notable difference in the size of the vocabulary item: 22% of the science vocabulary consisted of complex phrases, but none of the ELA took this form. These complex phrases in science were primarily two-word phrases (e.g., *chemical change*) but some were three or more words (e.g., *wheel and axle*). Exclusion of these items would have limited an understanding of the science vocabulary. At the same time, including words such as *change* in the phrase *chemical change* or *and* in *wheel and axle* might cause an underestimation of the difficulty of the vocabulary learning task in science. Consequently, the decision was made to analyze the rarer of the words in a phrase (e.g., *chemical* rather than *change* in *chemical change* and *wheel, axle* and not *and* in *wheel and axle*).

Seven features of the words (209 from the ELA program and 207 from the science program) were established, five of which have been used in numerous studies of vocabulary: (a) length of words (in letters); (b) predicted frequency per million words of text (Zeno, Ivens, Millard, & Duvvuri, 1995); (c) morphological frequency, that is, predicted frequency per million words of text of the words transparently related to the focus word, e.g., *revolve, revolving* for *revolution* but not *revolt* (Zeno et al., 1995); (d) familiarity based on the Living Word Vocabulary (Dale & O'Rourke, 1976) and its extension by Biemiller (2008); and (e) dispersion, which indicates how widely a word appears in different subject areas (Zeno et al., 1995). We use the space available in this chapter to describe the two features of focus: conceptual complexity and relatedness. Readers interested in more extensive descriptions of these variables are encouraged to examine the literature review provided by Scott, Lubliner, and Hiebert (2005).

With respect to conceptual complexity, Nagy et al. (1987) reported that a dichotomous grouping of their categories (1 through 3) versus 4 (highly complex) accounted for differences in readers' knowledge of vocabulary. After numerous iterations, we developed a three-point coding system. The definitions for the words that were provided by the publisher in either the teacher's guide (for the ELA words) or the glossary of the student book (for the science words) were entered into a database. The definitions were matched against the 2,000 words in the GSL (West, 1953). When a definition consisted of one or two words that were among the 2,000 most frequent words, the word was rated as 1 (the least complex). For example, *anticipation* was coded as 1 because

it was defined as "hope," which appears on the GSL. Words with definitions that were a single word that was not among the 2,000 most frequent words on the GSL were designated as category 2 (e.g., *quarantine* was defined as "isolation"). Where definitions consisted of phrases where all words were within the GSL, the word was also coded as 2 for conceptual complexity (e.g., "tool that measures wind speed" for *anemometer*). Definitions with phrases or clauses where at least one key word was not within the GSL were designated as the highest level of complexity. For example, *rotation* was defined as "the spinning of a planet, moon, or star around its axis." Because both *planet* and *axis* are not within the GSL, *rotation* was rated as having the highest level of complexity.

The measure for the relatedness feature drew on Marzano and Marzano's (1988) categorization of 7,300 words into 61 superclusters. After eliminating grammatical categories and consolidating several superclusters (e.g., Facial Expressions" with Communication), Hiebert (2011) identified 13 megaclusters that pertain to "big" ideas about story elements (e.g., Communication, Emotions & Attitudes) and the content of informational texts (e.g., Social Systems, Human Body). Whereas the original superclusters (Marzano & Marzano, 1988) were presented in order of size, Hiebert suggested that the vocabulary megaclusters be considered in three large groups: (a) words that would be expected to be distinctive of narrative vocabulary (e.g., Emotions & Attitudes, Character Traits), (b) words shared by both types of texts (e.g., Comparatives & Causes) and (c) words that are most prominent in informational texts (e.g., Natural Environment).

Results. Means and standard deviations for the measures, except for relatedness, are presented in Table 2. Results of statistical comparisons of features across the two sets of vocabularies are also included in Table 2. Differences were statistically significant for all of the measures except for the frequency of morphological families of words and the dispersion index.

The words in the narrative vocabulary are more likely to be familiar to students than the words in the science corpus but are predicted to appear less frequently. Although they are less familiar but more frequent, the science words are significantly longer and have definitions that are more conceptually complex

TABLE 2

Means and Standard Deviations for Features of Words in Narrative and Informational Texts

	Narrative	Informational	F (significance level)
Familiarity (LWV Grade)	6 (2.5)	7.5 (3.4)	42.752 (.000)
Frequency (U function)	13.7 (52.4)	39.1 (118.1)	28.039 (.000)
Frequency of Morphological Family	26.7 (116.4)	31 (78.4)	.275 (.600)
Dispersion Index	.60	.61	3.289 (.070)
Length	7.3	7.8	28.677 (.000)
Conceptual Complexity	1.4	2.3	275.941 (.000)

than the narrative set of words. The words in the narrative texts appear less frequently but, as evident in the findings on familiarity, students likely know their underlying meaning. The greater accessibility of the narrative vocabulary is evident in the conceptual complexity findings, which show a lower conceptual complexity rating for the narrative than for the science vocabulary.

Semantic relatedness was considered by examining the number of megaclusters represented within the target words for a unit of text (i.e., a story in the ELA program and a chapter in the science program). A ratio was developed for the average number of target words per instructional unit (7 in the ELA program, 11 in the science program) and the number of megaclusters represented in that group for an individual instructional unit. The ratio for ELA vocabulary was 7:5, and for the science vocabulary 11:4. A t-test indicated that the difference in the ratios was statistically significant ($t = 8.2$, $p = .000$). Most target words in an ELA unit did not come from closely related semantic clusters, whereas the vocabulary for an instructional science unit had at least several words from the same megacluster.

We were also interested in whether particular megaclusters were associated with particular text types. The percentages of the two vocabularies falling into the megaclusters are presented in Table 3. As was predicted (Hiebert, 2011), particular megaclusters such as Emotions & Attitudes and Character Traits were heavily represented in the ELA vocabulary but not in the science vocabulary. Both vocabularies had a substantial number of words within Natural Environment; this megacluster accounted for almost half of the words in the science vocabulary, but only about 20% of the words in the ELA vocabulary.

TABLE 3
Distribution of Megaclusters in Vocabularies of Two Types of Texts

Dominant/Shared Text Types	Megacluster	Narrative Text	Informational Text
Narrative Dominant	Emotions & Attitudes	.09	0
	Character Traits	.09	0
	Social Relationships	.02	.01
Narrative/Content Shared	Action & Motion	.12	.06
	Communication	.10	.09
	Characters	.10	.07
	Places Events	.06	.01
	Social Systems	.06	.01
	Physical Attributes (Objects, events, time)	.05	.08
	Comparatives & Causes	.03	.05
Content Dominant	Natural Environment	.19	.48
	Machines	.08	.07
	Human Body	.03	.07

An Analysis of the Features of Exemplar Texts

Although analyses of target words provide a view of what is taught or believed critical to teach, the representativeness of these words in relation to the entire corpus of words in texts also needs to be established if the vocabulary demands of texts are to be understood. To capture the nature of vocabulary in entire texts of the two text types, an exemplar was chosen from each program. The exemplars were the texts from which the two excerpts in Table 1 were taken. The ELA and the science text each came from the same place in its respective program—the third text of the third unit. For the ELA program, the text was *How Night Came From the Sea: A Story From Brazil* (Gerson, 1994) from Afflerbach et al. (2007). For the science text, the selection was "Why does matter have energy?" (Cooney et al., 2006). The former consisted of 1,250 words and the latter of 1,350 words.

Three features of the vocabulary within these two texts were of interest: (a) the ratio of different or unique words (also known as *types* in analyses of vocabulary) in relation to total words (typically referred to as *tokens* in vocabulary analyses), (b) the distribution of the unique and total words across different frequency groups, and (c) the number of repetitions of the targeted or assessed vocabulary within the texts. For the second feature, words were clustered into three groups based on the predictions of Zeno et al. (1995) for appearances of words per million words of text: (a) highly frequent words (appearances of 100 or more per million words), (b) moderately frequent words (appearances of 10–99 per million words), and (c) rare words (appearances of 9 or less per million words).

Results. Data summarized in Table 4 indicate that the ratio of unique to total words for the ELA and science exemplars was .33 and .26, respectively. The ELA text had substantially more unique words than the science text. The information in Table 4 also shows that twice as many of the unique words within the ELA text fell into the rare category than was the case with the science vocabulary. For readers to be proficient at reading the ELA text, they must have a considerably greater capacity to recognize unique words, either by already knowing the meaning of these words or by being able to extract the meanings from the context of the text.

TABLE 4

Distribution of Word Zones: Narrative and Informational Exemplars

Word Zones	Narrative Text			Informational Text		
	Total words (n=1,250)	Unique words (n=410)	Average # appearances	Total words (n=1350)	Unique words (n=328)	Average # appearances
Rare (WZ 5, 6)	.06	.15	1.2	.04	.03	5.4
Moderate (WZ3,4)	.14	.24	1.8	.27	.16	6.6
High (WZ0-2)	.79	.61	4.0	.69	.81	3.5

The number of appearances of words according to word zones is evident in Table 4. The patterns for words appearing with rare and moderate frequency differ substantially in the narrative and informational texts. Few of the rare words appeared more than once in the narrative text, while rare words in the informational texts appeared an average of five times. The pattern was the same for the words of moderate frequency, with substantially more appearances of these words in the informational than in the narrative texts. Within the informational text, students have the opportunity to become facile with the same word as it appears repeatedly in the text. For the narrative text, however, students must have the facility to understand many unique words that occur a single time in the text and that they are unlikely to have encountered in previous texts.

What Might These Differences in the Vocabularies of Narrative and Informational Texts Mean for Instruction?

The patterns from our study showed both quantitative and qualitative differences in the words identified for instruction with ELA and science texts. First, in comparison to the science text, the exemplar ELA text had more unique words, and more of these unique words were rare. The words called out for instruction accounted for 1% of the unique words in the ELA text. Another 14% of the unique words fell into the rare category of words that are unlikely to be encountered frequently in written language. By contrast, 3% of the words in the science text fell into this category, and with few exceptions, these words were the focus of instruction. Even within a text-based vocabulary effort, which the ELA program represents, instruction focuses on only a very small percentage of the words that are likely challenging for many students—especially for the two-thirds of an American fourth-grade cohort that reads at a less than proficient level (Daane, Campbell, Grigg, Goodman, & Oranje, 2005). Particularly in schools where the majority of students fall into this less-than-proficient group, teachers need to provide substantial scaffolding if students are to develop facility with vocabulary sufficient for comprehending narrative texts with any depth.

A second way in which the two exemplar texts differed was in the repetition of the targeted vocabulary. In addition to scaffolding students' recognition of the many words that fall outside the instructional focus, teachers need to do considerable scaffolding of the words chosen for instruction in the ELA text, because almost all of the instructional words appeared only once. Research is limited on the number of encounters that are required for a word to be known with any level of facility and precision (Swanborn & De Glopper, 1999). A single encounter with a word may be sufficient for learning to pronounce it (Share, 1995), but it is unlikely that a single encounter in a text will result in deep and generalizable understanding of a word's meaning. All of the words in a narrative text do not have to be known to get the gist of the action or dilemma. However, when narrative texts consume a large portion of the elemen-

tary curriculum, students may be exposed to many words, but they may not be expanding their facility with many of these words.

A third difference between the vocabularies of the two programs offers a potential solution for what may appear to be an insurmountable instructional challenge for teachers in ELA programs: The vast majority of the words called out for instruction in the ELA program (58%) were of the simplest conceptual complexity. Only 3% of the ELA vocabulary was of the highest level of conceptual complexity, and these words came from the limited number of informational texts that were part of the program. All but a handful of the words in the ELA program can be explained easily relative to students' existing concepts.

When that feature is combined with a fourth difference between the two vocabularies, a direction for instruction of the vocabularies of ELA texts that are primarily narrative becomes even more clear. The unique vocabularies in the two text types came from different vocabulary megaclusters. For the ELA texts, half of the words came from five clusters that have to do with characters—their names, traits, ways of communicating, actions and motions, and emotions and attitudes. Although the relatedness of words within an individual ELA story was limited, the connectedness across stories was substantial. This connectedness reflects the nature of narratives, not any concerted effort on the part of the publisher. The publisher does not give a rationale for the selection of particular words for particular stories. We suspect, however, that particular megaclusters would have been even more heavily populated had all of the unique, rare words for the stories within the ELA program, rather than the target vocabulary, been analyzed.

As Biber (1988) and other linguists have pointed out, authors of narrative and informational texts have different goals and, as a result, use words in very different ways. To underscore a theme in the story, Gerson (1994) in *How Night Came From the Sea* does not repeat any single word describing brightness, but she does repeat the concept of brightness with numerous different words (e.g., *shimmering, gleamed, brightness, brilliant, glittering*). By contrast, the authors of the science text (Cooney et al., 2006) repeat words such as *heat* and *radiation* numerous times. Cooney et al. are intent on developing a precise meaning of radiation and heat, but Gerson wants the reader to get a sense of the dilemma of the goddess's daughter, who longs for respite from the relentless sun. The characteristics of characters and contexts are repeated in the same narrative but with different words. With many different authors writing narratives that each contain many different words, the situation may seem insurmountable for ensuring that students understand the words that are used in a particular narrative. The task may seem to be a hopeless one when the goal is to build capacity so that students have the vocabulary to read complex narratives independently (Common Core State Standards Initiative, 2010). But there are similarities in the vocabulary of narrative texts that can be taught. Regardless of the narrative or an author's use of vocabulary, the same underlying concepts of traits, communication, features of contexts, and the nature of problems can

be expected to appear across narratives. When vocabulary instruction uses the words of particular texts to teach students to be cognizant of words used to teach the shared components of narratives, vocabulary learning can become generative.

We illustrate the nature of this instruction that is unique to narrative shortly. But before engaging in that discussion, it is important to understand why the nature of instruction for informational texts needs to be of a different kind than that for narrative texts. A text on the lifecycle of amphibians will contain words and descriptions that are unique, different from the words and descriptions in a text on the ways in which thermal energy is created. Authors of these texts will use different words as well as different text structures to communicate these constructs. But within a particular topic such as thermal energy, the same words are likely to appear again and again. In a subsequent grade when the topic reappears in a science text or in another book on the topic of thermal energy, students can expect that many of the same words will appear, such as *radiation, conduction,* and *convection*. These different purposes and their resulting different vocabularies suggest significantly different programs for instructional concepts and vocabulary in ELA and science. It would take a book-length manuscript to flesh out all of the details and uniquenesses of the vocabulary programs called for with different subject areas, but we outline here the main elements of these two types of vocabulary instruction.

Implications for the Instruction of the Vocabulary of Science Texts

We begin with two caveats about the vocabulary of science texts. First, although we explore what the differences in word features mean for instruction related to the texts that students read, we want to emphasize that we are not viewing the words of science texts as simply learned through vocabulary lessons. To understand radiant heat or convection requires numerous activities in addition to reading. In the Seeds of Science/Roots of Reading project where we have worked to integrate literacy and science content and instruction (Cervetti, Jaynes, & Hiebert, 2009), a four-part mantra guides the lessons: "Do it, talk it, read it, write it." Words such as *convection, conduction,* and *insulators* are used dozens of times in discussions, demonstrations, and writing activities. At least preliminary evidence suggests that such multimodal experiences appear to support the learning of conceptually complex words in science (Cervetti, Barber, Dorph, Pearson, & Goldschmidt, 2009).

A second caveat is that, because our analysis considered science texts only, conclusions cannot be generalized to other content areas such as social studies. A perusal of Marzano's (2004) summary of the vocabulary found in national and state standards suggests that two features that were associated with science vocabulary may be even more pronounced in social studies: complex phrases and polysemous words. Some of the observations that follow about these two features are likely to apply to social studies vocabulary, but we caution that this is a hypothesis only.

With respect to complex phrases in science, 22% of the words in the sample were accompanied by one or more words (*solar cell, solar energy, solar system*). Even when words function as a single idea, it is rare that these words are presented as compound words or even hyphenated to alert the reader to their concatenation. The complex phrase has a unique meaning that cannot necessarily be determined by understanding common meanings of each word individually. The presence of numerous complex phrases adds a challenge for students in reading science that needs to be addressed within instruction. This instruction is unlikely to occur if vocabulary is primarily emphasized in reading narratives. Only one of the words in the ELA vocabulary sample was a phrase (*boarding school*).
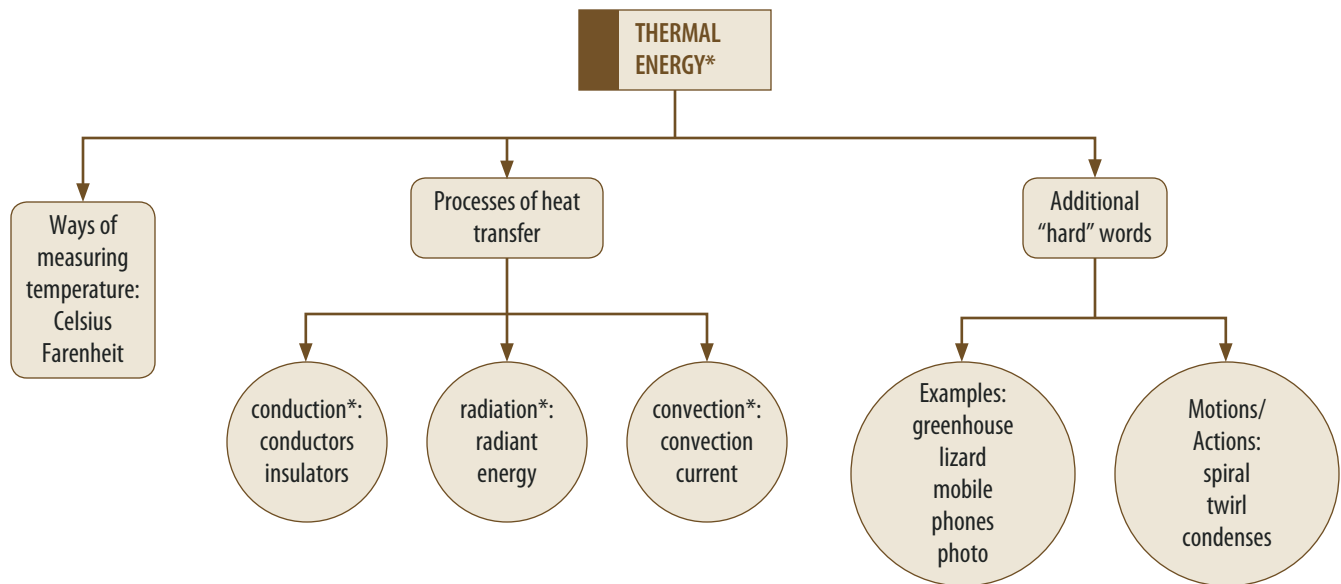
A second feature of the science vocabulary that has consequences for instruction was the higher average frequency rating of these words than for those in the ELA sample. The unique words in informational texts are often more frequent because they have multiple meanings across different subject areas. Many of the fundamental ideas within the science vocabulary—*work, speed, energy, force*—also have meanings that are used in everyday conversations. The word *work* has 53 common meanings according to Dictionary.com (http://dictionary.reference.com/). In the science program, one meaning only—and in this case a very precise one—is developed, which is work as "using force in order to move an object a certain distance" (Cooney et al., 2006, p. EM9). For both students and teachers, the ordinary, everyday meanings of such a word may mean that knowledge of the word is assumed. It is also the case that the everyday meanings of words that have popular meanings in nonscience contexts can interfere with students' understanding of the scientific meaning (Cervetti, Hiebert, & Pearson, 2010).

Critical distinctions in the meanings of scientific vocabulary will be made only through multiple forms of inquiry and discussion. Further, because the majority of science words represented conceptually complex ideas—even with ordinary labels such as *work, force, energy, speed, tissue, matter*—meanings need to be taught in relation to one another. A thematic map with the interrelationships of vocabulary is provided in Figure 1 to illustrate the connections among the complex ideas in the exemplar science text. The meaning of one conceptually complex word typically relies on an accurate (and precise) meaning of another conceptually complex word. These understandings are built through demonstrations, illustrations, DVDs, discussions, experiments, and writing. Everything in science cannot be experienced firsthand, but there are numerous ways in which background knowledge can be built through secondhand observation and inquiry.

The network of complex concepts also depends on experiences over time. The concepts in this unit (matter and thermal energy) were part of units in the primary grades. These concepts will be revisited in subsequent grades in even greater depth. If science is given short shrift in the primary grades, students will not develop the foundation for elaborations of existing concepts and new

Thematic Clustering of Unique, Rare Words within a Science Prototypical Text

```
                              THERMAL
                              ENERGY*
        ┌────────────────────────┼───────────────────────────┐
   Ways of               Processes of heat            Additional
  measuring                  transfer                "hard" words
 temperature:
   Celsius           ┌───────────┼───────────┐      ┌──────────┴──────────┐
  Farenheit      conduction*:   radiation*:   convection*:   Examples:        Motions/
                 conductors      radiant       convection    greenhouse       Actions:
                 insulators      energy        current        lizard           spiral
                                                               mobile           twirl
                                                               phones           condenses
                                                               photo
```

concepts that will be added to the thematic networks in higher grades. They will not have the capacity to read the increasingly more complex texts—a capacity which is the goal of the Common Core State Standards (Common Core State Standards Initiative, 2010).
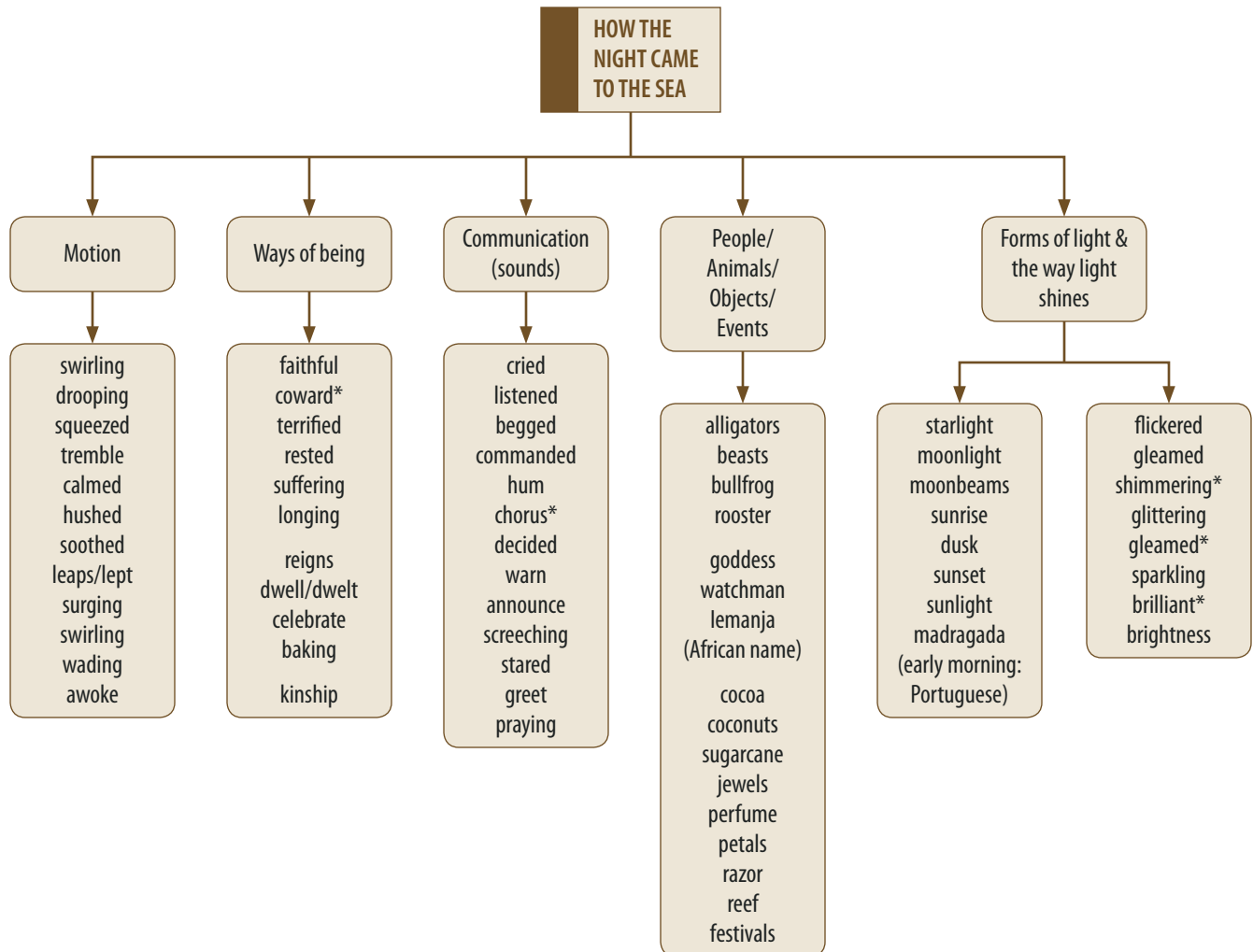
Implications for the Instruction of Words in Narrative Texts

The vocabulary of science is conceptually complex and requires intensive experiences over time; however, the vocabulary of the ELA program is dense with rare words. These rare words are typically not members of heavily populated morphological networks as is the case with the rare words in science (e.g., *shimmering* in the former; *nonrenewable* in the latter). They do not have the thematic connections within or across stories that characterize the words of the science curriculum. Where the core ELA program is *Houghton Mifflin Reading* (Cooper et al., 2004), vocabulary instruction for fourth graders focuses on *homage, commotion, hosted, severed,* and *fluffed* for a week; however, students in states or districts that have selected Scott Foresman's *Reading Street* (Afflerbach et al., 2007) are learning *chorus, coward, gleamed, shimmering,* and *brilliant.* From one program to another, there is little overlap (except in the few cases where the same story appears and even then target words can vary considerably). There is no rhyme or reason to selection of vocabulary within the ELA programs to which the lion's share of class time is devoted in American classrooms.

Nagy and Hiebert (2010) identified criteria for the selection of vocabulary within ELA programs. They underscored that, to close the vocabulary gap, the focus of instruction with narrative texts should be the unfamiliarity of words. This may sound like a strange criterion, but research over an extended period

FIGURE 2

Semantic Clustering of Unique, Rare Words within an ELA Prototypical Text

HOW THE NIGHT CAME TO THE SEA

Motion	Ways of being	Communication (sounds)	People/ Animals/ Objects/ Events	Forms of light & the way light shines

Motion
swirling
drooping
squeezed
tremble
calmed
hushed
soothed
leaps/lept
surging
swirling
wading
awoke

Ways of being
faithful
coward*
terrified
rested
suffering
longing

reigns
dwell/dwelt
celebrate
baking

kinship

Communication (sounds)
cried
listened
begged
commanded
hum
chorus*
decided
warn
announce
screeching
stared
greet
praying

People/Animals/Objects/Events
alligators
beasts
bullfrog
rooster

goddess
watchman
lemanja
(African name)

cocoa
coconuts
sugarcane
jewels
perfume
petals
razor
reef
festivals

Forms of light & the way light shines

starlight
moonlight
moonbeams
sunrise
dusk
sunset
sunlight
madragada
(early morning:
Portuguese)

flickered
gleamed
shimmering*
glittering
gleamed*
sparkling
brilliant*
brightness

of time suggests that students already know many of the words identified for instruction within basal reading programs. Almost 50 years ago, Gates (1962) demonstrated that the majority of words chosen for instruction in basal reading programs were already known sufficiently for students to comprehend the texts. More than 20 years ago, Stallman et al. (1990) confirmed the same pattern. Although we did not test students' understanding of the core vocabulary from a current core reading program (Afflerbach et al., 2007), 37% of the target vocabulary was rated as familiar for fourth graders (Biemiller, 2008; Dale & O'Rourke, 1976) and 60% of the words were ones that could be defined with a single word within the 2,000 most frequent words in written English (West, 1953).

A second criterion suggested by Nagy and Hiebert (2010) was that instruction of literary vocabulary emphasizes a metalinguistic perspective where groups of words and underlying linguistic features are the focus, rather than a word-by-word perspective. The exemplar text, *How Night Came From the Sea* (Gerson, 1994), is typical of narrative texts in that it has numerous words that belong to

rich semantic clusters. Nuanced words are used to convey how characters communicate, how they feel, and how they resolve their dilemmas and problems. Most fourth graders, even those who struggle as readers, have an understanding of basic concepts such as cowardice, yearning, fascination, and destruction, even though they may not use these words . All words cannot be taught, but readers can be taught to be aware that writers use multiple ways to label basic concepts about communications, feelings, traits, and settings. To expand vocabularies, students require the fundamental ideas of what stories are about and how writers of stories use rich vocabulary to communicate the human experiences. We propose that instructional scaffolds such as story structure and the cluster approach that have fallen by the wayside over the past two decades, are resources for both teachers and learners in developing richer vocabularies and more efficacious vocabulary instruction. In Figure 2, we have mapped out the numerous unique words in the exemplar text. Most words appeared a single time in the text and communicate nuances that readers require to grasp the style and gist of the text. When the words are viewed in relation to underlying concepts that cut across stories, however, numerous words can be addressed. Such an approach offers to expand students' vocabularies substantially more than the identification of seven or eight of the many unique words in the texts, most of which come from discrete vocabulary clusters.

Summary

In this chapter, we have illustrated that there are substantially different kinds of vocabularies offered in ELA versus science programs. These differences in vocabularies lend themselves to significantly unique instructional approaches. In science, most words are conceptually complex and represent new concepts for many students. These concepts are not learned by rote but evolve from extensive discussion, demonstrations, and experiments. The words that are unique to narrative texts are often numerous but represent concepts with which most students are familiar. Students may never have encountered the particular words that an author uses to convey a particular trait or motive of a character. It is likely, however, that even younger elementary students have underlying knowledge about the traits, motives, ways of moving, and emotions of characters. To become adept with narrative texts, students must understand the ways in which authors vary their language to ensure that readers grasp the critical features of the story. If the vocabulary gap is to be narrowed for the students whose academic learning experiences occur primarily in schools, educators need to develop unique selection criteria and instructional strategies for the vocabularies of both narrative and informational texts.

References

Afflerbach, P., Blachowicz, C. L. Z., Boyd, C. D., Cheyney, W., Juel, C., Kameʻenui, E. J., et al. (2007). *Reading street*. Glenview, IL: Scott Foresman.

Armbruster, B. B., & Nagy, W. E. (1992). Vocabulary in content area lessons. *The Reading Teacher*, *45*(7), 550–551.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biemiller, A. (2008). *Words worth teaching*. Columbus, OH: SRA/McGraw-Hill.

Cervetti, G. N., Barber, J., Dorph, R. Pearson, P. D., & Goldschmidt, P. G. (2009, April). *Integrating science and literacy: A value proposition?* Symposium paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Cervetti, G. N., Jaynes, C. A., & Hiebert, E. H. (2009). Increasing opportunities to acquire knowledge through reading. In E. H. Hiebert (Ed.), *Reading more, reading better* (pp. 3–29). New York, NY: Guilford.

Cervetti, G. N., Hiebert, E. H., & Pearson, P.D. (2010). *Factors that influence the difficulty of science words*. Santa Cruz, CA: TextProject, Inc.

Common Core State Standards Initiative. (2010). *Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects.* Washington, DC: Council of Chief State School Officers & National Governors Association.

Cooney, T., et al. (2006). *Scott Foresman Science*. Glenview, IL: Pearson.

Cooper, J. D., Pikulski, J. J., Ackerman, P. A., Au, K. H., Chard, D. J., Garcia, G. G., et al. (2004). *Houghton Mifflin Reading*. Boston: Houghton Mifflin Company.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2) 213–238.

Daane, M. C., Campbell, J. R., Grigg, W. S., Goodman, M. J., & Oranje, A. (2005). *Fourth-grade students reading aloud: NAEP 2002 special study of oral reading* (NCES 2006-469). U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. Washington, DC: Government Printing Office.

Dale, D., & O'Rourke, J. (1976). *The living word vocabulary*. Elgin, IL: Field Enterprises Educational Corporation.

Davis, F. B. (1942). Two new measures of reading ability. *Journal of Educational Psychology, 33*, 365–372.

Entwisle, D. R. (1966). *Word associations of young children.* Baltimore, MD: John Hopkins Press.

Fisher, C., Berliner, D., Filby, N., Marliave, R., Cahen, L., & Dishaw, M. (1980). Teaching behaviors, academic learning time, and student achievement: An overview. In C. Denham & A. Lieberman (Eds.), *Time to learn*. Washington, DC: National Institute of Education.

Gardner, D. (2004). Vocabulary input through extensive reading: A comparison of words found in children's narrative and informational reading materials. *Applied Linguistics, 25*(1), 1–37.

Gates, A. I. (1962). The word recognition ability and the reading vocabulary of second and third grade children. *The Reading Teacher*, *15*(6), 443–448.

Gerson, M. (1994). *How night came from the sea: A story from Brazil.* New York, NY: Little Brown & Co.

Hart, B. & Risley, T. (1995). *Meaningful differences in everyday parenting and intellectual development in young American children*. Baltimore, MD: Brookes.

Hiebert, E.H. (2011). *Growing capacity with the vocabulary of English language arts programs (Reading Research Report #11.02).* Santa Cruz, CA: TextProject, Inc.

Jenkins, J. R., & Dixon, R. (1983). Vocabulary learning. *Contemporary Educational Psychology, 8*(3), 237–260.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English based on The British National Corpus.* London: Longman.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–38.

Marzano, R. J. (2004). *Building background knowledge for academic achievement.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., & Marzano, J. S. (1988). *A cluster approach to elementary vocabulary instruction*. Newark, DE: International Reading Association.

Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen–Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, memory, and cognition, 21*(4), 863–883.

Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly, 19*(3), 304–330.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, *24,* 237–270.

Nagy, W., Berninger, V. W., & Abbott, R. (2006). Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle–school students. *Journal of Educational Psychology*, *98*(1), 134–147

Nagy, W. E., & Hiebert, E. H. (2010). Toward a theory of word selection. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. P. Afflerbach (Eds.), *Handbook of reading research* (Vol. 4, pp. 388–404). New York, NY: Longman.

Norris, S. P., Phillips, L. M., Smith, M. L., Baker, J. J., & Weber, A. C. (2008). Learning to read scientific text: Do elementary school commercial reading programs help? *Science Education*, *92*(5), 765–798.

Oxford Dictionaries. (2010). *Oxford Dictionary of English* (3rd Rev. ed.). New York, NY: Oxford University Press.

Scott, J. A., Lubliner, S., & Hiebert, E. H. (2005). Constructs underlying word selection and assessment tasks in the archival research on vocabulary instruction. In J. V. Hoffman, D. L. Schallert, C. M. Fairbanks, J. Worthy, & B. Maloch (Eds.), *55th Yearbook of the National Reading Conference* (pp. 264–275). Oak Creek, WI: NRC.

Sereno, J., & Jongman, A. (1997). Processing of English inflectional morphology. *Memory and Cognition, 25*, 425–437.

Share, D. L. (1995). Phonological recoding and self–teaching: *sine qua non* of reading acquisition. *Cognition*, *55*(92), 151–218.

Simpson, J., & Weiner, E. (2009). *Oxford English Dictionary*. New York, NY: Oxford University Press.

Stallman, A. C., Commeyras, M., Kerr, B. M., Meyer-Reimer, K., Jiménez, R., Hartman, D. K., & Pearson, P. D. (1990). Are "new" words really new? *Reading Research and Instruction, 29*(2), 12–29.

Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. (1979). Memory representations for morphologically related words. *Journal of Verbal Learning and Verbal Behavior, 18,* 399–412.

Sternberg, R., & Powell, J. S. (1983). Comprehending verbal comprehension. *American Psychologist, 38,* 878–893.

Swanborn, M. S. L., & De Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research, 69*(3), 261–285.

Thorndike, R. L. (1973). *Reading comprehension education in fifteen countries*. Stockholm, Sweden: Almquist & Wiksell.

Tinkham, T. (1993). The effect of semantic clustering on the learning of second language vocabulary. *System, 21*, 371–380.

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research*, *13*(2) 138–163.

Waring, R. (1997). The negative effects of learning words in semantic sets: A replication. *System, 25*, 261–274.

West, M. (1953). *A General Service List of English Words*. London, UK: Longman.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide.* New York, NY: Touchstone Applied Science Associates.

TextProject, Inc. is a non-profit public benefit corporation. Its aim is to bring beginning and struggling readers (of any age) to high levels of literacy through a variety of strategies and tools, particularly the texts used for reading instruction.

Find out more at **textproject.org**