CrossMark

Applied Psycholinguistics **36** (2015), 1441–1457
doi:10.1017/S0142716414000344

A graphomorphemic approach to identifying and selecting a set of high utility, stable affixes common to the technical vocabulary of science

KRISTIN NELLENBACH, JENNIFER ZOSKI, JOY DIAMOND, and KAREN ERICKSON
University of North Carolina at Chapel Hill

Received: October 22, 2013 Accepted for publication: April 5, 2014

ADDRESS FOR CORRESPONDENCE
Kristin Nellenbach, Department of Allied Health Sciences, Center for Literacy and Disability Studies, University of North Carolina at Chapel Hill, Room 1110 Bondurant Hall, 321 South Columbia Street, Chapel Hill, NC 27599–7335. E-mail: kristin_nellenbach@med.unc.edu

ABSTRACT
Adolescents often learn science vocabulary through reading. This vocabulary is frequently characterized by multisyllabic words derived from Greek and Latin roots. While most adolescents have acquired the decoding skills to read these multisyllabic words, many students, particularly those with disabilities, cannot engage in independent word learning because they lack the skills to decode these multisyllabic words. Graphomorphemic elements of words, including affixes, support effective decoding and can eventually support word learning. This article describes an approach used to identify the most frequently occurring, stable affixes within science words so that they could be used in "big word" decoding instruction. To illustrate the approach, a subset of high frequency science words and a list of high utility, stable affixes are provided.

There is growing national concern regarding the performance of our nation's adolescents on assessments of science literacy (i.e., National Assessment of Educational Progress, 2006; National Research Council, 2007, 2010). A driving force behind this concern is that a majority of our nation's youth, armed with less than proficient skills in science literacy, will be unable to successfully compete in the global economy. Because a portion of science instruction uses text content as a primary source of information, students' comprehension of such written information is essential to acquiring mastery of the content. However, the construction, load, and/or integration of science vocabulary within secondary science text can present a substantial challenge for some readers, leading some researchers to suggest that the "language of science, rather than content, is a major barrier (if not *the* major barrier) to most pupils in learning science" (Wellington & Osborne,

© Cambridge University Press 2014 0142-7164/14 $15.00

https://doi.org/10.1017/S0142716414000344 Published online by Cambridge University Press

2001, p. 2). Further compounding the challenge is that some students are unable to successfully decode complex, multisyllabic science words in order to begin to access their meanings (Lowe, 2011).

Most teachers of adolescents rightly assume that students acquire the ability to decode multisyllabic words in the elementary grades. For example, in the Common Core State Standards–English Language Arts, students are expected to decode multisyllabic words by fifth grade (see e.g., Reading Foundation Standard 5.3a). Unfortunately, some students have not acquired these skills. As a result, teachers are not only faced with the challenge of teaching the meanings of words that students encounter in text, but also find themselves having to support the development of decoding skills that are required to read the words for more effective access to their meanings. With an overabundance of academic and discipline-specific multisyllabic words embedded in science materials, science teachers in particular may find effective and direct instruction of such words to be a daunting if not impossible task. Therefore, it is important to provide instruction that teaches students the skills required to independently decode novel and difficult multisyllabic words in the science texts they come across each day so that they may eventually access the meaning of those words independently.

Developing effective advanced decoding instruction depends on knowing which constituent parts to teach. Given that more than 90% of multisyllabic words in science and technology are derived from Greek and/or Latin roots (Green, 2008) and that those roots contribute to lexical complexity in numerous ways (Bar-Ilan & Berman, 2007), these specific constituent parts are especially important. A consolidated list of these parts could be used to guide science teachers, support professionals, and others in their judicious selection of important graphomorphemic elements and to inform subsequent decoding instruction. To contribute to a growing body of research in this area, this article describes a process for generating a list of multisyllabic science words and a graphomorphemic approach to identifying such a consolidated list of the most frequently occurring, stable affixes within those words.

LITERATURE REVIEW

The role of multisyllabic words in science

Multisyllabic or big words are long words that consist of eight or more letters, contain two or more syllables, and in many cases, multiple morphemes (Cunningham, 1998). Although adolescents must learn to read and understand large numbers of new, multisyllabic words across domains (Archer, Gleason, & Vachon, 2003; Curtis, 2004; Kamil, 2003), the density of the multisyllabic words within science texts is especially high (Fang, 2005, 2008). Many of the big words students encounter in science text are derived from Greek and/or Latin roots (Fang, 2006; Green, 2008) and present numerous levels of challenge relative to their conceptual difficulty, number of alternate meanings, and specialized meaning in science (Hiebert & Cervetti, 2012). Successful decoding, spelling, and comprehension of these more sophisticated word forms requires knowledge and skills that are different from

https://doi.org/10.1017/S0142716414000344 Published online by Cambridge University Press

those learned in the primary grades when first learning how to read (Carnegie Council on Advancing Adolescent Literacy, 2010; Fang, 2008; Lowe, 2011).

In the primary grades, young children are introduced to and explicitly taught how to read, spell, and pronounce hundreds of simply constructed, one- or two-syllable words (National Institute of Child Health and Human Development, 2000). Most of these words occur in oral and written language and are generalized across the curriculum rather than associated with a particular domain. Explicit decoding instruction in the primary grades emphasizes teaching students to attend first to individual letters and their corresponding sounds and then to the spelling patterns in phonograms and syllable types. Younger students may also be introduced to and taught simple morphological word forms, such as common roots and affixes (prefixes and suffixes), but the big words they encounter in the upper grades, particularly the technical vocabulary of science, require much more attention to these forms. The systematic study of the spelling–sound connection (i.e., letter–sound correspondences and basic syllable patterns) that students have learned to rely on in order to read, spell, and comprehend simple words in the primary grades does little to support students in reading the complex, technical vocabulary they encounter in secondary science texts (Green, 2008; Lowe, 2011).

In comparison with the primary grades, the multisyllabic words adolescents encounter are more complex because many of the word forms are new, occur at a much lower frequency, and are more domain specific (Beck, McKeown, & Kucan, 2008). According to triple word form theory, when children learn to read they must attend to multiple linguistic areas (i.e., the orthographic, phonological, and morphological word forms and their parts), and they must coordinate this knowledge to successfully and efficiently decode words (Berninger et al., 2006). While simpler word forms tend to emphasize the connection between orthography and pronunciation, there appears to be a stronger underlying link between orthography and meaning in complex, multisyllabic words. This seems to be especially true for the technical vocabulary of science because most of the multisyllabic words in science have spelling patterns with Greek or Latin origins (Green, 2008), which not only convey precise concepts and understandings specific to the sciences (Shanahan & Shanahan, 2012) but also help to support and preserve word meanings, even in the event of changes in pronunciation (Cunningham, 1998; Templeton, 2012). If adolescents are going to be successful in reading and ultimately comprehending the multisyllabic technical words of science, explicit instructional approaches that support their independent recognition of the spelling–meaning based patterns within these big words is critical.

Instructional approaches to decoding multisyllabic words

The research addressing the most effective methods of teaching adolescent readers to decode multisyllabic words points in two directions: (a) syllabification strategies (e.g., Bhattacharya & Ehri, 2004; Shefelbine, 1990) and (b) graphomorphemic strategies that include deliberate exposure to high-frequency orthographic and morphological word elements (e.g., Carlisle & Stone, 2005; Cunningham, 1998). Each has its benefits as an approach to decoding, but the graphomorphemic

https://doi.org/10.1017/S0142716414000344 Published online by Cambridge University Press

approach offers the added benefit of providing additional access to the meaning of the word once it has been decoded (Cunningham, 1998; Templeton, 1991).

Syllabification approach. Skilled readers can, with automaticity, analyze the vowel and consonant patterns within big words to determine how to segment and pronounce them because of their knowledge of frequently occurring spelling patterns and interletter frequencies (Adams, 1990). Weak associations develop between letters that are rarely seen together, thus enforcing syllable boundaries. By fourth grade, competent readers are able to perceive syllable units more quickly and accurately than single letters (Adams, 1990). There is some evidence to suggest that syllabication, the process of analyzing the vowel and consonant patterns within big words to determine how to segment and pronounce them, can lead to improvements in general word reading skills (Bhattacharya & Ehri, 2004; Shefelbine, 1990) as well as word reading within the content areas (Bhattacharya, 2006). As suggested by Bhattacharya (2006), this type of syllabification strategy could be extremely helpful for teaching older struggling readers because it can be incorporated into content-area lessons across the curriculum.

However, there are a few caveats in teaching adolescents to break apart big words using a syllabication approach. The first caveat is that in many multisyllabic words there is more than one acceptable way to segment words into syllables. For example, the word *carnivore* has three syllables and could be broken apart as *car-ni-vore* based on syllable boundary rules, yet could also be acceptably segmented as *car-niv-ore* and *carn-iv-ore.* So while use of syllabication strategies have been shown to improve decoding of multisyllabic words (see Battacharya & Ehri, 2004; Shefelbine, 1990), flexibility in syllable parsing is recommended for successful decoding interventions. The second caveat relates to the variability of orthographic units when parsing multisyllabic words into syllables using pronunciation rules. For example, analyzing the vowel and consonant patterns within biology to determine its pronunciation results in the following segmentation, *bi-ol-o*-gy, but a pronunciation-based segmentation of *biological* can logically result in the following division *bi-o-log-i-cal.* Further, this approach leads to segmenting *microwave* as *mi-cro-wave, and micrometer* as *mi-crom-et-er.* These examples illustrate that while syllabification helps with pronunciation, it creates a variety of orthographic patterns. Notice that *log* is an orthographic unit in *biological,* but is completely separated between syllables in *biology* and *cro* in *microwave* becomes *crom* in *micrometer.*

In addition to the orthographic inconsistencies that result from this approach to syllabification, the resulting syllables, unlike morphemes, do not necessarily carry or relate to a word's meaning and therefore do not support access to the meaning of the word once it has been decoded. Using the examples above, it is only the first syllable in the third iteration, *carn*, that is the graphomorphemic element related to the word's identifiable meaning, flesh, whereas *car* does not. In addition, in the second and third syllabication iterations, *ore* does not refer to eating, yet *–vore* is the meaning-based element from the Latin word *vorare*, meaning to eat, swallow, or devour (Green, 2008; McCarthy, 2012). Similarly, *ol* in *biology* and *crom* in *micrometer* do not carry meaning. Further, while a syllabication approach can help students pronounce multisyllabic words, it does not offer the same generative

https://doi.org/10.1017/S0142716414000344 Published online by Cambridge University Press

access in the way that a graphomorphemic approach to decoding does. Recognition of the syllable chunk *mi* in the word *microorganism*, for example, does not offer the same efficient generative power and access to hundreds of other unknown words as does recognition of the larger graphomorphemic element *micro.*

In summary, while syllabication can help students access the pronunciation of words, it fails to produce consistent orthographic units or provide access to the meanings of words in the way that a graphomorphemic approach to decoding does. In order to emphasize the graphomorphemic connection within words, we need to teach students to be able to recognize the meaning-based parts and their associated orthographic patterns so that *carnivore*, becomes segmented as *carn-(i)/vore*; *biology* becomes *bio-logy; biological* becomes *bio-logi-cal, microwave* becomes *micro-wave* and *micrometer* becomes *micro-meter*. Like Carlisle and Stone (2005), we support the view that morphemes, particularly in the context of science vocabulary, play an integral role in decoding these complex words using an approach that supports access to the meaning of those words once they have been successfully decoded.

Graphomorphemic approach. Similar to students' use of syllable knowledge for syllabication, skilled readers may also use their knowledge of morphemes, the smallest units of meanings in words, to read unfamiliar big words encountered in text. Morphological knowledge includes awareness of base words (which may include roots), inflections (which change the tense, case, or person of a base), and derivations (which change the word class and/or meaning of a base). For older students, such knowledge is a better predictor of decoding skills than are phonological skills (Mann & Singson, 2003). Further, it has been posited that sensitivity to the morphemic structure of words plays a key role not only in word reading (Carlisle & Stone, 2005; Henry, 1989; McCutchen, Logan, & Biangardi-Orpe, 2009) but quite possibly also in word learning (McCutchen & Logan, 2011).

Because many of the words secondary students encounter in science are multisyllabic (and multimorphemic) derived word forms, it is reasonable to advocate for an approach that teaches students to attend to these forms for independent word reading. Such an approach, referred to as a graphomorphemic approach, uses morphemes in much the same way that phonemes are used in beginning phonics instruction to teach advanced decoding skills for big word reading. Long advocated by several researchers (e.g., Cunningham, 1998; Henry, 1989; Templeton, 1991), a graphomorphemic approach teaches students to attend to the larger meaning-based chunks (i.e., morphemes) within words *and* their corresponding, underlying spelling–meaning and pronunciation patterns. Cunningham (1998) forwarded the idea that explicit, direct instruction of a set of high-utility affixes could provide struggling readers with a foundational storage of multisyllabic words that they could read, spell, and associate with meaning, leading to a widely used word list for beginning big word decoding instruction. This word list, known as the "nifty thrifty fifty" (Cunningham & Hall, 1998), is a list of 50 words that includes the prefixes, suffixes, and spelling changes that are the most prevalent in the multisyllabic words that intermediate students may encounter. Henry (1989) found that a graphomorphemic approach targeting high-frequency, consistently spelled

https://doi.org/10.1017/S0142716414000344 Published online by Cambridge University Press

prefixes, suffixes, and bases (together called graphomorphemic elements) leads to significant gains in intermediate students' decoding and spelling abilities.

Several approaches to sound graphomorphemic instruction are described in the literature (e.g., Abbott & Berninger, 1999; Cunningham & Hall, 1998; Reed, 2008; Templeton, Bear, Invernizzi, & Johnston, 2010). We favor an approach that provides students with repeated exposure to a variety of deriving and decomposing activities embedded with a series of sequenced, integrated lessons. We also support an approach that provides students with ample opportunities to practice reading, spelling, and using targeted affixes within meaningful contexts. Activities such as identifying, sorting, deriving, and decomposing multisyllabic words into their constituent parts directs students' attention to the important morphologic, orthographic, and phonetic patterns and teaches them the skills necessary for independent big word decoding. Using a word building activity for example, students can be taught how to derive multiple, new word forms given a selected bank of affixes and bases whereby the students manipulate them to create words. While the specifics may vary, each of these graphomorphemic approaches teaches students a set of high-utility morphemes and then teaches them to use these morphemes to decode and spell by deriving and decomposing real words that contain the high utility morphemes.

Whether students are taught to decode multisyllabic words using an approach to decoding that is based on syllables or graphomorphemic elements, successful decoding of multisyllabic words requires attention to phonology, orthography, and morphology (Berninger et al., 2006). As Berninger et al. (2003) argue, it is not the task of older readers to attend to the grapheme–phoneme correspondences underlying the alphabetic principle; instead, students in the upper elementary grades and beyond must learn to attend to the connections between larger meaning units and the printed word. Thus, morphology and its interrelationship with orthography and phonology appear to play a key role in the process by which older readers learn to accurately and efficiently decode complex multisyllabic words. Therefore, we also forward the notion that direct instruction of frequently occurring morphemes (common to science words) with attention to phonology, orthography, and morphology can lead to improvements in advanced decoding as well as support word learning.

Importance of high-frequency, stable morphemes to decoding

The ability to recognize high-frequency base words and affixes supports decoding of lower frequency multisyllabic words (Carlisle & Stone, 2005; Goodwin, Gilbert, & Cho, 2013). Successful word decoding requires quick and efficient access to words and/or morphemes. When students read high-frequency words, they decode them as whole-word forms, with speed of access determined by whole-word frequency (Alegre & Gordon, 1999). However, multisyllabic science words are often low frequency and domain specific. Thus the frequencies of parts of these words (i.e., morphemes) are essential in determining quick and efficient access to the whole word (McCutchen et al., 2009). Several recent studies have addressed this specific issue. Carlisle and Stone (2005), for example, found that accessing the base morphemes when reading can facilitate access to morphologically

https://doi.org/10.1017/S0142716414000344 Published online by Cambridge University Press

complex words. Deacon, Whalen, and Kirby (2011) found that children were faster at reading derived words that have higher base morpheme frequencies, reinforcing the claim that children are more efficient at decoding multimorphemic words when they are composed of high-frequency word elements or morphemes. These results also support Carlisle and Stone's (2005) finding that the frequency of morphemes contributes to reading accuracy for low-frequency words. Overall, data from this line of research strengthens the assertion that reading speed and accuracy are influenced by the frequency of the morphemes that students encounter when reading.

In addition to frequency, the orthographic and phonological consistency or stability of morphemes has also been suggested to support decoding of complex, multisyllabic words. Orthographic consistency generally refers to the degree to which letters or letter clusters map onto sounds. Studies have indicated that children more easily learn to read words that are orthographically consistent or stable than words that contain inconsistent spelling–sound patterns (Georgiou, Parrila, & Liao, C., 2008; Yap & Balota, 2009). However, English spelling patterns are not very consistent at the individual phoneme–grapheme level, but they become more consistent when the vowel is considered with consonant units, and both children and adults use these vowel–consonant units to support their decoding (Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995). In addition, in many multisyllabic science words, phonological shifts (i.e., changes in pronunciation) occur between a base word and a derived multisyllabic word, particularly with the addition of suffixes, resulting in inconsistent meaning–sound connections. However, as Templeton suggests, "orthography clarifies, where pronunciation may obscure, relationships among words" (2012, p. 117). Therefore, in cases when students are attempting to access complex multisyllabic words, consideration of the underlying connection between orthography and morphology must come into play as "spelling visually preserves these meaning or morphological relationships among words" (Templeton, 2012, p. 117).

Given that there seems to be a consistent or stable underlying connection between orthography and meaning within many technical big words in science, it is also important to consider the semantic stability of specific graphomorphemic elements, to most effectively choose morphemes that will help students decode, spell, and eventually comprehend big words (Templeton, 2012). Consistent with Perfetti's (2007) lexical quality hypothesis, students may be better able to form high-quality lexical representations, allowing them to efficiently decode words, when they develop strong bonds between the semantic and orthographic (as well as phonological) components of words. Morphology may play a key role in the development of these high-quality lexical representations such that teaching students to recognize morphological units will speed their access to words, freeing up resources for higher level reading comprehension. Exposing students to affixes with consistent or stable spelling–meaning links across words may help them begin to develop an awareness and recognition of these larger linguistic units, lay a pathway for initial access to words via decoding, and ultimately reinforce the underlying meaning connections necessary for supporting growth in comprehension.

Although both syllabification and morpheme-based interventions have been shown to improve multisyllabic decoding, the morpheme-based approach has

https://doi.org/10.1017/S0142716414000344 Published online by Cambridge University Press

the added benefit of supporting vocabulary learning in addition to improving decoding. In addition, the frequency of the morphemes that students encounter along with their associated consistency or stability patterns is suggested to be integral to supporting students' ability to independently and efficiently read, spell, and comprehend complex, multisyllabic words. Knowledge of the most frequently occurring, consistent or stable morphemes within the technical words of science can be an invaluable tool to science teachers and related professionals who wish to support those students who struggle with big word decoding. In the next section, we describe how a graphomorphemic approach was used to guide our selection of a set of high-frequency, stable and flexible affixes extracted from a developed list of science words.

METHODS

There is a solid base of research supporting the use of a graphomorphemic approach for teaching word reading (Cunningham, 1998; Templeton, 1983, 2012; Templeton et al., 2010; White, Power, & White, 1989). Furthermore, there are clearly identified lists of the most common prefixes, suffixes and base words (roots) in the corpus of words in written English (e.g., Cunningham, 1998; Henry, 1990, 2010; White, Sowell, & Yanagihara, 1989) for use in a graphomorphemic approach to teaching students to decode multisyllabic words. However, few morpheme lists exist to guide the use of a graphomorphemic instructional approach within the domain of science (for an exception see Templeton, Bear, Invernizzi, Johnston, 2010), and to date, there is no known accessible resource representative of the words found in grades 6–12 science materials in the United States from which frequently occurring morphemes could be selected. As such, it was necessary to first develop a science-specific word list in order to determine the frequently occurring morphemes within the words on that list.

Developing a science-specific word list for Grades 6–12

A few well-established corpora were initially explored for accessibility and guidance in the development of our science-specific word list. First, two international corpora were explored, including the Academic Word List (AWL; Coxhead, 2000) and the British National Corpus (BNC, BNC XML Edition; BNC Consortium, 2007). The AWL was developed by Coxhead (2000) in New Zealand to guide teachers in identifying important vocabulary to teach students in postsecondary education. It is an academic-based corpus compiled from over 3 million running words of written English found among texts across four different subject areas, including arts, commerce, law, and science (Coxhead, 2000). Because science words were included in the AWL, headwords of the word families, along with words from a related science-specific word list (Coxhead & Hirsh, 2007) were used in the development of our word list. The other international corpus, the BNC, is a collection of 100 million words drawn from a collection of samples of written and spoken language representing a wide cross-section of British English in the later part of the 20th century (http://www.natcorp.ox.ac.uk/corpus/index.xml). Ninety million of the words in the BNC are from written English. The BNC

https://doi.org/10.1017/S0142716414000344 Published online by Cambridge University Press

include tags that enable a search resulting in a list of all words in the corpus found in science-related texts. Given the tagging accessibility of the BNC, we used the science words identified from this corpus as a way to cross-reference our eventual smaller, science-specific word list.

Another important corpus available in the United States, the Corpus of Contemporary American English (COCA, Davies, 2008), was also considered. This corpus consists of 450 million words, of which 22.8 million words are tagged as science and technology specific words; however, the science and technology specific words within the COCA are extracted from peer-reviewed scholarly articles and academic-oriented magazines rather than the texts encountered by students in Grades 6–12 in the United States. Furthermore, the COCA cannot be searched in a manner that allows the extraction of all words tagged from science sources. For these reasons we were not able to use the COCA as a resource in the development of our word list. However, because the COCA does allow the end-user to search specific words to determine the frequency of the words within the COCA corpus, the frequency of words included in our developed list were checked in the COCA so that their importance beyond Grade 12 could be considered when selecting them for instruction.

Next, a sampling of other accessible resources that contained words from across the various science domains (i.e., chemistry, earth/space science, life science, and physics) consistently taught in public schools across the United States was collected. The collection of resources included journal articles, state-generated vocabulary lists, dictionaries, educator-created resources, and a standards-based document for a total of 15 resources (see online only Supplementary Materials Appendix S.1). A majority of the compiled resources included identifiable lists of words; these words were simply extracted and cataloged within a database that was created using Microsoft Excel. In addition to adding general documentation descriptors, such as resource codes, words, and syllable counts, when information was available, domain area and grade-level notations were added. The standards-based document, *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (National Research Council, 2012), was particularly valuable in its use as a singular, comprehensive source of representation across the science domains. Although this document does not include specific word lists, words reflective of core and component ideas/content were embedded throughout the grade band endpoints within the Framework Dimensions sections that students are exposed to in texts and classroom materials. Several noun and verb word forms from the Framework Dimension sections were added to the database, such as *chemical*, *molecules*, *interacting*, *reactants*, and *properties*.

Our initial compilation of science-specific words resulted in 7,274 entries that included word repetitions and monosyllablic as well as multisyllabic words. After all repetitions and words with only one or two or syllables were removed, the resulting list included 3,642 words. A subset of frequently occurring words appearing in at least two or more of the consulted resources were identified from within our word list, resulting in a count of 830 high-frequency science words (830 HFSW), with the complete listing available in online only Supplementary Materials Appendix S.2. The resulting word list was determined to be reasonable enough in size and representation to reliably reflect the types of multisyllabic

https://doi.org/10.1017/S0142716414000344 Published online by Cambridge University Press

science words adolescents encounter such that we could next engage in the process of morpheme selection.

Generating a list of high-frequency affixes common in 6th–12th grade science words

Using the 830 HFSW list, the authors independently hand-selected graphomorphemic elements within the words, compared results, and resolved any interrater disagreements. Due to variability in parsing, however, outside sources were consulted for further refinement and reliability in the selection process. Unable to locate a single comprehensive listing of morphemes in line with the research objectives, a list of approximately 1,500 science-based morphemes was compiled (see online only Supplementary Materials Appendix S.3 for resources). To assess the validity of this listing, a computer program was developed to search for the frequency of each of the 1,500 morphemes across the larger corpus of science-specific words extracted from the BNC. Each of the 1,500 morphemes and their corresponding set of matched words were initially reviewed, if applicable, across all three positions of words including the beginning (i.e., prefix), middle (i.e., base), and end (i.e., suffix). Prefixes were defined as graphomorphemic elements occurring in the initial positions of words, and suffixes were defined as elements occurring in the final positions of words. We quickly recognized that identifying and selecting morphemes in the middle positions of words using this method was not particularly useful. For example, the morpheme */eco/* returned words like, ***deco***mpose and *milli**seco**nd*. Therefore, the subsequent analysis of the 1,500 morpheme list was restricted to reviewing only those morphemes that occurred in the beginning or ends (i.e., affixes) of words. Once the affix process was completed we would then be able to, at a future date, best determine how to proceed with identifying and selecting base morphemes.

The initial effort to identify the frequency of affixes from the list of 1,500 morphemes resulted in observed inflated counts, especially within the BNC, most likely attributable to its being a much larger corpus. For example, the morpheme */on/* was included in word counts of longer morphemes, like *con-, -ion,* and *mono-* leading to multiple counts for the morpheme */on/*. We also noted inconsistent counts related to variations in spelling depending on the morpheme's adjacent letters and position in words, like *gene* in ***gen****otype* and *ion* in *convec**tion***. To address these inconsistencies, a second computer-based analysis was conducted in which morphemes were searched from longest to shortest so that once a longer morpheme was identified, a shorter one found within it would not be counted. For example, if a word contained the morpheme *-tion*, then the morphemes *-ion* and *-on* would not be identified for that word. In addition, through multiple iterations of scanning the resulting morpheme list and the corresponding words, we added acceptable spelling variations of some morphemes to our list in order to capture patterns of orthographic changes within our morphemes, like *gen* in *genotype* versus *gene* in *genetic*. We were thus able to extract all spelling variants of morphemes within our word list. We then chose the spelling variant that was most prevalent among these words in order to prioritize high-frequency science morphemes with stable spelling–meaning connections for beginning instruction. In the example

https://doi.org/10.1017/S0142716414000344 Published online by Cambridge University Press

Table 1. *Example of a process to determine the stability of high utility affixes*

Morpheme	Science Words	Phonologically Stable	Semantically Stable
meso-	Mesoderm, mesoglea, mesophyll, Mesozoic	4	4
hypo-	Hypothermia, hypothesis, hypotonic, hypoxia	0	2
-scope	Gyroscope, microscope, spectroscope, stethoscope, telescope	4	4
-ance	Dominance, inheritance, reactance, resistance, resonance	4	0

Note: 4 = 100% stability patterns, 2 = 51%–99% stability patterns, 0 = ≤50% stability patterns.

of *gen/gene*, we chose *gen,* as it allowed us to capture this morpheme in words that followed the spelling patterns of both of these examples. Through choosing only the most frequent and/or representative spelling variant per morpheme, we were essentially able to control for the consistency or stability of the orthographic patterns within the morphemes up front, aiding in our computer-based analyses. This secondary analysis resulted in a narrower set of affixes that was then subsequently run against our list of 830 HFSW to determine a set of the most frequently occurring affixes with stable or consistent spelling patterns. This analysis resulted in a finalized selection of 111 affixes (63 prefixes and 48 suffixes). All affixes were then assigned a frequency score from 2 to 10, in increments of 2, based on the number of times the affix occurred within our 830 HFSW word list. After having identified and selected a set of high-frequency, orthographically stable affixes, we began an iterative process of further refining the list based on additional word characteristics.

Identifying and selecting stable affixes

Having generated a list of high-utility affixes common to the sciences, the next step in the process was to review this list for the particular word characteristics known to support multisyllabic word decoding, namely consistency or stability. Stability refers to the degree by which the affixes were consistent in either their phonological pronunciation or meaning across our set of 830 HFSW. First, determining phonological stability required ensuring that an individual affix was pronounced the same in each word. Based on the authors' collective clinical judgment, each affix was coded in relation to a set of words in which it occurred (see Table 1 for an example). Affixes that were pronounced consistently with minimal or no changes were scored higher (i.e., 4) than those that were not pronounced consistently (i.e., 2 or 0). Any sounds that fell within a range of acceptable pronunciations across all dialects of American English or included vowel changes due to stress alone were given a score of 4. Using the same approach, the second step involved determining

https://doi.org/10.1017/S0142716414000344 Published online by Cambridge University Press

the semantic or meaning-based stability of each affix across the set of words in which it appeared. Again, this involved the authors' review to determine if the meaning of the affix was stable or consistent and easily accessible across the set of words in which it appeared. In other words, affixes that were both stable in their meaning as well as being accessible or transparent (i.e., knowledge of the meaning of the affix provides support for accessing the meaning of the word) were scored higher than those with unstable meaning patterns or that were opaque (i.e., knowledge of the affix provides little or no support for accessing the meaning of the word). Semantically transparent affixes provide readers with a clue for accessing the meaning of the full word when they have read or heard the affix in a related word before. For example, students are likely to know the word *electricity*, and thus recognize the meaning of the morpheme *electr-* in other words. Therefore, affixes with consistent, accessible meanings were scored as high (i.e., 4) while those with moderate to frequent changes in meaning were low (i.e., 2 or 0).

It is important to note that all scoring decisions were determined based on the orthographic, phonological, and semantic patterns that were consistent within our list of 830 HFSW. Words outside of the 830 HFSW list will have different spelling, pronunciation, and meaning patterns than those identified and selected here. Science teachers and related professionals should be aware that these patterns often change, particularly with derived words forms, and they will have to go through a similar process with their own list of words.

Weighted composite score

Having clinically judged and scored our set of affixes in relation to frequency, phonological stability, and semantic stability, we decided to weight the list of affixes as a final step in our approach. The decision to create a weighted list was based on our alignment with the literature and in keeping with our overall goal to identify a set of affixes that can be used in an instructional approach to develop big word decoding skills. Thus the weighting process was conducted in order to determine priorities for instruction. We wanted to choose the most stable affixes for early instruction and save less consistent affixes for later lessons. In this way, students would still need to use flexible decoding strategies, but they would be exposed to more consistent opportunities early in instruction. For each of the three characteristics (frequency, phonological stability, and semantic stability) a normalized score from 0 to 1 was generated. Next, to create the weighted composite score the three normalized scores were summed using associated weights (*frequency* = 1, *phonological stability* = 1, and s*emantic stability* = 0.5), and the final composite score was normalized to a scale from 0 to 100. Because our main goal is to help students first access these words through decoding, we prioritized frequency and phonological stability over semantic stability at this point. However, all of the affixes were originally chosen as morphemic units and therefore the spelling–meaning relationship was preserved. The resulting lists of weighted affixes (see online only Supplementary Materials Appendixes S.4 and S.5) are an important starting place in planning a graphomorphemic instructional program to teach students to decode multisyllabic science words.

https://doi.org/10.1017/S0142716414000344 Published online by Cambridge University Press

DISCUSSION

Instructional and clinical implications

The selection process and guiding graphomorphemic approach described in this paper builds on a growing body of research regarding multisyllabic word decoding. It resulted in two resources that can be used with relative confidence as preliminary tools for providing direct instruction in advanced decoding and augmenting vocabulary instruction. Science teachers and support personnel may find that the lists provide important guidance regarding which affixes to emphasize in decoding instruction and in their vocabulary instruction. In this way, science teachers can help students not only learn the words that are targeted in a particular class or unit but also learn to attend to the affixes within those words that have the greatest utility in decoding other science words.

A similar approach to the one described in this manuscript could also be followed to identify the high-utility morphemes in words in other content areas even at the level of a single textbook or unit of study. With the increasing availability of books in digitized formats, it is feasible for an individual teacher or team of teachers to conduct frequency counts of words that appear in the texts they use, sort those words by length, and parse the multisyllabic words to identify the morphemes within those words. Within a collaborative framework, teachers can use this process by identifying and providing discipline-specific word lists to specialists (e.g., speech–language pathologists) who in turn can lend their expertise in language to assist teachers in guiding the process of identifying and selecting high-utility morphemes with the linguistic characteristics known to best support big word decoding.

The lists that resulted from the process described in this manuscript are currently being used in the development of software that will teach adolescents to use a graphomorphemic approach to decoding multisyllabic science words. Eventually, additional software will be created that will automatically generate decoding instructional content from textbooks that teachers upload. In the meantime, educators can use the results of this effort to improve the impact of their decoding and/or vocabulary instruction related to science.

This study is not without limitations, but it does represent a systematic approach to identifying and prioritizing affixes for instruction targeting big word decoding. Teaching students a strategy for looking for these elements in words and using them as a bridge to pronunciation is the first step in eventually supporting students in accessing the meaning of the words they read. In the future, it will be important to continue building on and refining the 830 HFSW that resulted from the approach described in this manuscript. It would be useful to incorporate multisyllabic words that appear in frequently used textbooks in middle and high schools across the United States and to add to the larger corpus of words using tools such as the COCA (Davies, 2008) when and if additional search features are added to it. It will also be useful to validate the parsing of words and weighting system and evaluate the impact of graphomorphemic approaches to decoding instruction using the prioritized affixes.

https://doi.org/10.1017/S0142716414000344 Published online by Cambridge University Press

Conclusion

The technical vocabulary of science, characterized by complex, multisyllabic word forms, poses an initial challenge to word learning for some students because they lack advanced decoding skills. Teaching students to decode big science words can take a couple of forms, but a graphomorphemic approach offers the most promise because it provides access to the spelling and pronunciation of the word as well as the meaning (Cunningham, 1998). A graphomorphemic approach also assists in developing a shared understanding of morphemes that have the highest utility. Prior to the work described in this paper, there was no known empirically derived list of morphemes that students are most likely to encounter in science text content across Grades 6–12. The efforts described here resulted in such a list. The list of 830 HFSW and affixes will be used by the authors to create a computer-based instructional program, but it can also provide an important starting place for educators working to meet the needs of their students and other developers of curriculum and software.

SUPPLEMENTARY MATERIALS

The Supplementary Materials appendices can be found online at http://journals.cambridge.org/aps

ACKNOWLEDGEMENTS

The content of article is related to work the authors are conducting as part of a grant awarded to the University of North Carolina at Chapel Hill by the U.S. Department of Education, Office of Special Education Programs (Big Words II, H327A110023). The views expressed herein are solely those of the authors, and no official endorsement by the U.S. Department of Education should be inferred. The authors extend their gratitude to the two anonymous reviewers for their helpful comments on an earlier version of this manuscript.

REFERENCES

Abbott, S. P., & Berninger, V. M. (1999). It's never too late to remediate: Teaching word recognition to students with reading disabilities in grades 4–7. *Annals of Dyslexia*, *49*, 223–250.

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Alegre, M., & Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language*, *40*, 41–61.

Archer, A. L., Gleason, M. M., & Vachon, V. L. (2003). Decoding and fluency: Foundation skills for struggling older readers. *Learning Disability Quarterly*, *26*, 89–101.

Bar-Ilan, L., & Berman, R. A. (2007). Developing register differentiation: The Latinate–Germanic divide in English. *Linguistics*, *45*, 1–35.

Beck, I. L., McKeown, M. G., & Kucan, L. (2008). *Creating robust vocabulary: Frequently asked questions and extended examples.* New York: Guilford Press.

Berninger, V. W., Abbott, R. D., Thomson, J., Wagner, R., Swanson, H. L., Wijsman, E. M., et al. (2006). Modeling phonological core deficits within a working memory architecture in children and adults with developmental dyslexia. *Scientific Studies of Reading*, *10*, 165–198.

https://doi.org/10.1017/S0142716414000344 Published online by Cambridge University Press

Berninger, V. W., Nagy, W., Carlisle, J., Thomson, J., Hoffer, D., Abbott, S. et al. (2003). Effective treatment for dyslexics in grades 4–6: Behavioral and brain evidence. In B. R. Foorman (Ed.), *Preventing and remediating reading difficulties: Bringing science to scale*. Timonium, MD:York Press.

Bhattacharya, A. (2006). Syllable-based reading strategy for mastery of scientific information. *Remedial and Special Education*, *27*, 116–123.

Bhattacharya, A., & Ehri, L. C. (2004). Graphosyllabic analysis helps adolescent struggling readers read and spell words. *Journal of Learning Disabilities*, *37*, 331–348.

BNC Consortium. (2007). *British National Corpus (Version 3, XML ed.).* Distributed by Oxford University Computing Services on behalf of the BNC Consortium. http://www.natcorp.ox.ac.uk/

Carlisle, J. F., & Stone, C. A. (2005). Exploring the role of morphemes in word reading. *Reading Research Quarterly*, *40*, 428–449.

Carnegie Council on Advancing Adolescent Literacy. (2010). *Time to act: An agenda for advancing adolescent literacy for college and career success*. New York: Carnegie Corporation of New York.

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

Coxhead, A., & Hirsch, D. (2007). A pilot science-specific word list. *Revue francasie de linguistique appliquee*, *12*, 65–78.

Cunningham, P. M. (1998). The multisyllabic word dilemma: Helping students build meaning, spell, and read "big" words. *Reading & Writing Quarterly*, *14*, 189–218.

Cunningham, P. M., & Hall, , (1998). *Month-by-month phonics for the upper grades*. Greensboro, NC: Carson-Dellosa.

Curtis, M. E. (2004). Adolescents who struggle with word identification: Research and practice. In T. L. Jetton & J. A. Dole (Eds.), *Adolescent literacy research and practice* (pp. 119–134). New York: Guilford Press.

Davies, M. (2008). *The corpus of contemporary American English: 450 million words, 1990–present.* Retrieved from http://corpus.byu.edu/coca/

Deacon, S. H., Whalen, R., & Kirby, J. (2011). Do children see the *danger* in *dangerous*? Grade 4, 6, and 8 children's reading of morphologically complex words. *Applied Psycholinguistics*, *32*, 467–481.

Fang, Z. (2005). *Scientific literacy: A functional linguistic perspective. Science Education*, *89*, 335–347.

Fang, Z. (2006). The language demands of science reading in middle school. *International Journal of Science Education*, *28*, 491–520.

Fang, Z. (2008). Going beyond the Fab Five: Helping students cope with the unique linguistic challenges of expository reading in intermediate grades. *Journal of Adolescent and Adult Literacy*, *51*, 476–487.

Georgiou, G., Parrila, R., Liao, C. (2008). Rapid naming speed and reading across languages that vary in orthographic consistency. *Reading and Writing*, *21*, 885–903.

Goodwin, A. P., Gilbert, J. K., & Cho, S-J. (2013). Morphological contributions to adolescent word reading: An item response approach. *Reading Research Quarterly*, *48*, 39–60. doi: 10.1002/rrq.037

Green, T. M. (2008). *The Greek and Latin roots of English*. Lanham, MD: Rowman and Littlefield.

Henry, M. K. (1989). Children's word structure knowledge: Implications for decoding and spelling instruction. *Reading and Writing*, *1*, 135–152.

Henry, M. K. (1990). *Words*. Austin, TX: Pro-Ed.

Henry, M. K. (2010). *Unlocking literacy: Effective decoding and spelling instruction* (2nd ed.). Baltimore, MD: Brookes.

Hiebert, E. H., & Cervetti, G. N. (2012). What differences in narrative and information texts mean for the learning and instruction of vocabulary. In E. J. Kame'enui & J. F. Baumann (Eds.), *Vocabulary instruction: Research to practice* (2nd ed., pp. 322–344). New York: Guilford Press.

https://doi.org/10.1017/S0142716414000344 Published online by Cambridge University Press

Kamil, M. (2003). *Adolescents and literacy: Reading for the 21st century*. Washington, DC: Alliance for Excellent Education.

Lowe, C. (2011, Winter). Top10 reasons for Latin. *Classical Teacher*. Retrieved October 18, 2012, from http://www.memoriapress.com/articles/Top10-Reasons-for-Latin.html

Mann, V. A., & Singson, M. (2003). Linking morphology knowledge to English decoding ability: Large effects of little suffixes. In E. Assink & D. Sandra (Eds.), *Reading complex words: Cross-language studies* (pp. 1–25). New York: Kluwer Academic.

McCarthy, E. (2012, August 1). *Macroevolution.net.* Online dictionary of biological and medical roots, prefixes, and suffixes. Retrieved from http://www.macroevolution.net

McCutchen, D., & Logan, B. (2011). Inside incidental word learning: Children's strategic use of morphological information to infer word meanings. *Reading Research Quarterly*, *46*, 334–339.

McCutchen, D., Logan, B., & Biangardi-Orpe, U. (2009). Making meaning: Children's sensitivity to morphological information during word reading. *Reading Research Quarterly*, *44*, 360–376.

National Assessment of Educational Progress. (2006). *The nation's report card: Science 2005*. Retrieved October 8, 2011, from the National Center for Educational Statistics at http://nces.ed.gov/nationsreportcard/pdf/main2005/2006466.pdf

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Retrieved from http://www.nichd.nih.gov/publications/nrp/smallbook.htm

National Research Council. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Washington, DC: National Academies Press. Retrieved October 5, 2012, from http://www.nap.edu/catalog.php?record_id = 11463

National Research Council. (2010). *Rising above the gathering storm, revisited: Rapidly approaching Category 5*. Washington, DC: National Academies Press. Retrieved October 5, 2012, from http://www.nap.edu/catalog.php?record_id = 12999

National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*, 257–383.

Reed, D. (2008). A synthesis of morphology interventions and effects on reading outcomes for students in grades K–12. *Learning Disabilities Research and Practice*, *23*, 36–49.

Shanahan, T., & Shanahan, C. (2012). What is disciplinary literacy and why does it matter? *Topics in Language Disorders*, *32*, 7–18.

Shefelbine, J. (1990). A syllabic-unit approach to teaching decoding of polysyllabic words to fourth- and sixth-grade disabled readers. In J. Zutell & S. McCormick (Eds.), *Literacy theory and research: Analysis from multiple paradigms* (pp. 223–230). Chicago: National Reading Conference.

Templeton, S. (1983). Using the spelling/meaning connection to develop word knowledge in older students. *Journal of Reading*, *27*, 8–15.

Templeton, S. (1991). Teaching and learning the English spelling system: Reconceptualizing method and purpose. *Elementary School Journal*, *92*, 185–201.

Templeton, S. (2012). The vocabulary–spelling connection and generative instruction. In E. J. Kame'enui & J. F. Baumann (Eds.), *Vocabulary instruction: Research to practice* (pp. 116–138). New York: Guilford Press.

Templeton, S., Bear, D. R., Invernizzi, M., & Johnston, F. (2010). *Vocabulary their way: Word study with middle and secondary students.* Boston: Pearson.

Treiman, R., Mullennix, J., Bijeljac-Babic, R. & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, *124*, 107–136.

https://doi.org/10.1017/S0142716414000344 Published online by Cambridge University Press

Wellington, J., & Osborne, J. (2001). *Language and literacy in science education*. Philadelphia, PA: Open University Press.
White, T. G., Power, M. A., & White, S. (1989). Morphological analysis: Implications for teaching and understanding vocabulary growth. *Reading Research Quarterly*, *24*, 283–304.
White, T. G., Sowell, J., & Yanagihara, A. (1989). Teaching elementary students to use word part cues. *Reading Teacher*, *42*, 302–308.
Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, *60*, 502–529.

https://doi.org/10.1017/S0142716414000344 Published online by Cambridge University Press