# Vocabulary Input through Extensive Reading: A Comparison of Words Found in Children's Narrative and Expository Reading Materials

DEE GARDNER

Brigham Young University

The role of extensive reading in building vocabulary continues to receive considerable attention in first and second language research and pedagogy. This study analyses the lexical differences between narrative and expository reading materials used in upper-elementary education (10- and 11-year-old children), and explores how these differences could affect children's potential vocabulary acquisition through reading. Results of a computerized analysis of nearly 1.5 million word tokens reveals marked differences between 28 narrative and 28 expository children's books in terms of overall token distribution and individual type repetitions at all levels of vocabulary analysed in the study (i.e. general high frequency words, academic high frequency words, and specialized words). Further exploration of the lexical data indicates high numbers of register-specific words at all levels of vocabulary, particularly at the more specialized levels where the potential for protracted vocabulary growth is the greatest. A subsequent discussion addresses qualitative differences in the characteristics of these exclusive narrative and expository types. These lexical findings are used to assess claims of *Wide Reading* and *Free Reading* relative to children's acquisition of vocabulary through extensive reading, especially the default claims of 'incidental' word acquisition through repetitive encounters with unknown words while reading large volumes of material for pleasure.

First language (L1) researchers investigating the English vocabularies of elementary school students (ages 5 through 12) and the vocabulary load of their texts have concluded that the average L1 student acquires approximately 3,000 new words each year (Miller and Gildea 1987; Nagy and Herman 1987; Nagy, Herman, and Anderson 1985), and that the bulk of the new acquisitions, particularly after the third grade (ages 8 and 9), come through natural reading experiences (Nagy 1988; Nagy and Herman 1987). Additionally, the 3,000 per year growth figure has led to the conclusion that the average 12th grader (ages 17 and 18) possesses a reading vocabulary of approximately 40,000 words (Nagy and Herman 1987). If one considers vocabulary items such as idioms, compounds, and derivatives, as well as multiple word meanings for the same word form, these figures may be as high as 4,000 to 6,000 annually, or approximately 80,000 by the 12th grade (Anderson 1996; Anderson and Nagy 1992).

The major premise behind these lexical determinations is that readers acquire incremental knowledge of new words over time through multiple exposures to those words in multiple and varied contexts. In this way, the 'sheer volume' of natural reading experiences makes up for the relatively inefficient process of obtaining word meanings from context, which itself is assumed to be the major means of children's vocabulary acquisition (Nagy 1988: 30). This default explanation of protracted vocabulary growth through the school years has come to be known as the *Incidental Acquisition Hypothesis*—a position which has had considerable influence on reading and vocabulary research during the past 20 years.

The major pedagogical conclusion emerging from this line of thinking is that educators should allow more opportunities for students to read self-selected texts during class time (*Free Reading*) and encourage them to read widely (*Wide Reading*) both inside the classroom and at home. Proponents have also argued that less time should be spent on direct vocabulary instruction in school (i.e. one word at a time) which could not possibly account for the phenomenal vocabulary growth noted above: 'Our results strongly suggest that a most effective way to produce large-scale vocabulary growth is through an activity that is all too often interrupted in the process of reading instruction: Reading' (Nagy, Herman, and Anderson 1985, p. 252—cf. Nagy and Anderson 1984; Nagy, Anderson, and Herman 1987; Nagy and Herman 1984, 1985, 1987).

During this same 20 year time frame, Krashen's 'Input Hypothesis' (1982, 1985) has gained substantial acceptance in many second language (L2) instructional settings, including L2 elementary education. The major tenets of this position are that true language acquisition is a subconscious process that takes place as a motivated and relaxed learner is exposed to natural language that is slightly above his or her current level of ability or knowledge ('i + 1'). In addition, this stance assumes that conscious attempts to gain language knowledge through overt instruction and focus on form rarely, if ever, lead to true acquisition.

As this view began to find its way into explanations of L2 reading development, proponents were quick to explicate the natural links between the Input Hypothesis and the *Incidental Acquisition Hypothesis* proffered by Nagy and associates in L1 settings, particularly regarding long-term vocabulary growth through natural reading experiences involving authentic and motivating materials (Cho and Krashen 1994; Krashen 1989, 1993a, 1993b; Elley 1991, 1996; Elley and Mangubhai 1983):

> According to IH [Input Hypothesis], when the Language Acquisition Device is involved, language is *subconsciously* acquired—while you are acquiring, you don't know you are acquiring; your conscious focus is on the message, not form. Thus, the acquisition process is *identical to what has been termed 'incidental learning'*. Also, acquired knowledge is represented subconsciously in the brain—it is what Chomsky has termed 'tacit knowledge.' (Krashen 1989: 440, emphases added)

Additionally, Whole Language and Krashenist views have been linked together on these same issues (Coady 1997a; Crandall 1993; Elley 1991). The associations of 'natural,' 'authentic,' and 'incidental' (without direct attention to form) coupled with the practices of 'free' and 'pleasureful' reading have therefore crossed and crisscrossed the L1 and L2 reading domains over the past 20 years.[1] Also, the corollary arguments against direct vocabulary instruction and overt attention to word learning forwarded by Krashen and whole language proponents have had a marked impact on pedagogical thinking in both settings (see Coady 1997b and Zimmerman 1997, for reviews):

> The results of *incidental* studies suggest that comprehensible input alone can do the *entire job for vocabulary* and nearly the entire job for spelling. . . . This is, I think, good news for readers. Many people, I am sure, would not read at all if they were compelled to *work on their spelling and vocabulary* while trying to enjoy a *good book*. (Krashen 1989: 448, emphases added)

Until recently the connections in the research literature between the *Input Hypothesis* in L2 and the *Incidental Acquisition Hypothesis* in L1 appear to have been made primarily from the L2 side. However, in more recent publications by Anderson (1996) and Nagy (1997), the researchers not only link their vocabulary research to Krashen's claims, but they also adopt the term 'comprehensible input,' a term made famous by Krashen, to describe the ideal condition for language acquisition.

  Despite the popularity of these similar philosophies, several L1 and L2 reading and vocabulary experts continue to question the default nature of the *Incidental Acquisition Hypothesis*, arguing that claims of large-scale vocabulary gains through sheer volume of reading experiences may de-emphasize the types of materials that readers are expected to read, the types of words that various materials expose readers to, and the degree to which readers may rely on sources outside of incidental context exposure in their word learning efforts (see, for example, Carrell and Carson 1997; Carver 1994; Carver and Leibert 1995; Kinsella 1997; Raptis 1997). In short, some experts find it disconcerting that most of the vocabulary-growth claims involving *Wide Reading* and *Free Reading* hinge on children's 'playful, stimulating experiences with good books' (Anderson 1996: 74). Such materials, these sceptics argue, may not be representative of the more lexically-dense, content-area materials that students are required to negotiate in order to be successful in their academic endeavors—a concern which, ironically, appears to be justified in incidental acquisition research itself:

> We found small but highly reliable increments in word knowledge attributable to reading at all grades and ability levels. The overall likelihood ranged from better than 1 in 10 when children were reading *easy narratives* to near zero when they were reading *difficult expositions*. (Anderson 1996: 61, emphases added)

While this finding certainly adds fuel to the 'good-stories' position taken by proponents of incidental vocabulary acquisition, it also raises some serious theoretical quandaries. For instance, what if the words to be learned in the easy narratives are not the same as the words in the difficult expositions? By extension, if some words are found only in the difficult expositions, can extensive gains in vocabulary knowledge during the school years be largely attributable to a mental process (i.e. subconscious incidental acquisition) that relies on language conditions not normally associated with the reading of such materials (i.e. 'playful', 'easy', 'i + 1')?

## DISTRIBUTION OF VOCABULARY IN WRITTEN MATERIALS

What is known to date about the distribution of vocabulary in 'authentic' written materials raises additional questions not easily answered by the default claims of the *Incidental Acquisition Hypothesis*.

For example, Nation (1990) insists that the lexical coverage of most adult-level academic texts written in English can be characterized according to the breakdown in Table 1 (cf. Nation and Newton 1997). These figures indicate that 2,000 word families[2] cover roughly 87 per cent of most academic texts, 800 subtechnical word families, known as the 'University Word List' (UWL), cover an additional 8 per cent, with the balance of text coverage being accounted for by technical word families (3 per cent) and low frequency word families (2 per cent). In essence, Nation is claiming that 95 per cent of the lexical coverage in most texts is accounted for by approximately 3,000 word families (High Frequency plus UWL),[3] with technical and low-frequency word families combining for the remaining 5 per cent of text coverage. However, the claim that only 800 word families cover an additional 8 per cent of academic texts is still being debated. For example, a study of Dutch university texts suggests that approximately 10,000 high frequency words are needed to approach the 95 per cent level (Hazenberg and Hulstijn 1996). While Nation (2001) continues to maintain that only 570 words (the Coxhead Academic Word List) covers 8.5 per cent of academic texts, he has greatly modified his high frequency word coverage for such texts to 78.1 per cent. This would mean that high frequency plus academic high frequency words cover only 86.6 per cent of the words in academic texts rather than the 95 per cent suggested previously.

Interestingly, several studies have also concluded that the threshold of vocabulary knowledge for adequate reading comprehension is around 95 per cent of the words in a given text (Hu and Nation 2000; Laufer 1989, 1997; Liu and Nation 1985; Wixson and Lipson 1991). This means that basic reading comprehension may require readers to know at least 19 of every 20 words they encounter in a given text, and that the meaning of an unknown word is more easily ascertained from context if approximately 19 of its neighbouring words are known—that is, that a minimum of 95 per cent of the words in a text are already known to the reader (Laufer 1997; Liu and Nation 1985;

*Table 1: Word types and text coverage*

|  | No. of words | Proportion of text (%) |
|---|---|---|
| High-frequency words | 2,000 | 87 |
| University word list | 800 | 8 |
| Technical words | 2,000 | 3 |
| Low-frequency words | 123,200 | 2 |
| Totals | 128,000 | 100 |

Adapted from Nation, 1990: 16

Nation 1993). While researchers may disagree on the approximate number of words that constitute 95 per cent of most texts, there appears to be less disagreement on the 95 per cent comprehension threshold itself.[4] Subscribers to the coverage position use such figures to justify direct, intensive instruction of the 2,000 or 3,000 or 5,000 most frequent words of English so that L1 and L2 readers will be able to recognize and negotiate those words automatically, thus freeing up their mental resources for higher-level reading comprehension and the learning of additional vocabulary in context (Coady 1993, 1997a, 1997b; Coady *et al*. 1993; Laufer 1997; Nation 1990; Nation and Newton 1997). Such thinking has also resulted in simplified text designs for reading development such as graded readers (Wodinsky and Nation 1988) and other types of vocabulary control which attempt to provide the learner with exposure to lexical items in gradual increments based on overall word utility (i.e. high frequency, high range, and high coverage).[5] (See Coady 1993, and Zimmerman 1997 for reviews.)

While there is general consensus that high-frequency words must be mastered in order to achieve minimum levels of reading proficiency in both L1 and L2, there is also agreement that all students must develop knowledge of technical or topical vocabulary in order to negotiate the specialized texts and subject matter they will encounter in academic settings. In fact, Grabe (1991) argues that, in specialized academic settings, infrequent words may be the most important for L2 readers to know.

The paradox found in Table 1 is that approximately 125,000 word families (i.e. Nation's conservative estimate of the bulk of the English word stock), represent only 5 per cent of the words actually used in texts. This means that, in general, most words appear very infrequently in written materials. Yet, in certain specialized contexts, they are often highly salient (Scarcella and Zimmerman 1998), and are largely responsible for the richness and precision of written expression: 'Technical vocabulary has a very narrow range; that is, it is used within a specialized field. Within that field it may be reasonably

common. It is likely that every field has its technical vocabulary or the equivalent (Nation and Newton 1997: 239–40).

The implication for extensive academic vocabulary growth is that after a basic level of high frequency word knowledge, the type of materials that students read will have a profound effect in determining which new words they are exposed to and the degree to which those words are recycled. Indeed, if a reader reads 'widely' from an assortment of unrelated texts, the chances of some important unknown words recycling over time could be greatly diminished. This dilemma can only be exacerbated by the fact that 'authentic' literature (no attempt to control language) is viewed as an essential component of *Wide* and *Free Reading*.

Thus, when extensive reading is defined as *Wide Reading*, it becomes, in one sense, an open-ended construct that makes no real distinction between reading related books and unrelated books, or between reading narratives and expository materials, and therefore rests on the somewhat shaky premise that unknown words, beyond a basic level of high frequency words, will actually be recycled in a reasonable time frame.

> Essentially, vocabulary learning from extensive reading is very fragile. If the small amount of learning of a word is not soon reinforced by another meeting, then that learning will be lost. It is thus critically important in an extensive reading program that learners have the opportunity to keep meeting words they have met before. (Nation 1997: 15)

## PRINT EXPOSURE STUDIES

In allying themselves to claims by Nagy and associates, Stanovich and his colleagues have also built a strong case for linking extended vocabulary growth to amount of print exposure rather than amount of speech exposure:

> There are sound theoretical reasons for believing that print exposure is a particularly efficacious way of expanding a child's vocabulary. These reasons derive from the differences in the statistical distributions of words found between print and oral language. (Stanovich *et al.* 1996: 20)

In their treatment of this topic, the researchers cite a study (Hayes and Ahrens 1988) which reported large disparities between the lexical compositions of oral speech samples and written text samples in terms of the number of rare word occurrences in each. The findings led Stanovich *et al.* to conclude that 'conversation is not a substitute for reading' when it comes to vocabulary growth, especially 'after the middle grades' (1996: 21).

However, from the perspective of the current study, these researchers may have overlooked an equally important issue—namely, that certain types of reading materials may not be adequate substitutes for other types of reading materials in exposing readers to 'rare' words (i.e. words that are not typically

found in high frequency lists), or that rare types may not be shared between different types of reading materials. An example of this potential oversight is the lack of discussion by the researchers regarding notable disparities between the vocabulary in the children's books, where the average word ranked 627 on the frequency list of the American Heritage Intermediate Corpus,[6] and the vocabulary in the popular magazines, where the average word ranked 1,399 on that same list.

While the differences between magazines and children's books may be largely accounted for by intended audience (i.e. adults vs. children), it is none the less perplexing why vocabulary differences in materials *within* the same broad register (i.e. oral register only or written register only) are not considered. Why, for instance, is the more lexically-rich courtroom testimony, labelled a 'special situation'(p. 20), not compared with the other oral registers of prime-time adult television and conversation of college graduates? Would lexical conclusions regarding the assumed speech-writing distinction have been different if courtroom testimony, assumed by these researchers to be an uncharacteristic speech register, were replaced by the more frequently experienced registers of news broadcasts or classroom lectures? By extension, would it make a difference if children's books were analysed in terms of narrative fiction versus expository nonfiction, or a history textbook versus historical fiction? In short, should the bulk of children's growth in knowledge of rare words be categorically attributed to reading experience in general without consideration of (a) differences in potential vocabulary exposure depending on the *type of written material* the children read? and (b) the possibility that some oral registers with higher informational characteristics may play a more prominent role in exposing children to important words than previously thought? (See Ellis and He 1999; Ellis, Tanaka, and Yamazaki 1994, and Gibbons 1998 for treatments of this latter issue in L2 settings.)

Indeed, Biber (1988) clearly cautions against drawing conclusions about language factors based on simple speech–writing dichotomies:

> There is no single dimension of orality versus literacy. That is, even the notions of 'oral' and 'literate' texts, taken to represent typical speech and writing, are multi-dimensional constructs . . . it is only when we restrict our comparison to conversation and academic exposition that we observe a single set of genre relations. (Biber 1988: 162–3)

Of particular importance to the current study is Biber's finding that varieties of fiction (i.e. romance, general, mystery, adventure, and science) exhibit many more speech-like characteristics than their academic-prose counterparts (cf. Hirsh and Nation 1992).[7] Nation and Waring's (1997) consolidation of several word-list studies supports these conclusions and points to global vocabulary distinctions between fictional and more academically-oriented reading materials in terms of both lexical density (i.e. novels are less lexically dense than academic material) and the presence of specific-purpose vocabulary (i.e. more such types appear in academic materials).

Taken together, these register-related findings suggest that the reading of narrative fiction may only account for a certain level or type of vocabulary growth:

> From a content perspective, narratives typically deal with information about social or interpersonal relationships and everyday problem solving, content about which both adults and children tend to know quite a bit. . . . Narratives have a fairly limited set of important relations among events, primarily causal and temporal. In contrast, a common purpose of expository texts is informational. Informational texts frequently present concepts and relations that readers do not already know. They require that readers understand a greater range of logical relations among pieces of information. (Coté, Goldman, and Saul 1998: 6)

Interestingly, in the study just cited, the L1 fourth and sixth graders verbally reported having the greatest difficulty in 'understanding the meaning of individual words or sentences' during their reading of informational texts (p. 21). Again, the nature of the reading materials that children are exposed to appears to have a profound effect on their overall reading experiences and vocabulary learning opportunities.

These convergent findings from seemingly diverse fields of inquiry provide linguistically-viable evidence to begin questioning certain default vocabulary claims of the *Incidental Acquisition Hypothesis* (cf. Raptis 1997). In short, one of the major problems appears to be that such claims may have understated the 'what' in children's reading experiences in an effort to emphasize the 'how much'. While the latter continues to enjoy almost universal acceptance, the former appears to warrant further scrutiny. One of the purposes of the current study is to more fully explore vocabulary variation between narrative and expository texts and the potential that such variation may have on large-scale vocabulary growth during the school years.

To date, it appears that research involving incidental vocabulary acquisition through extensive reading has been primarily concerned with estimations of large scale vocabulary gains through reading, determinations of the role of context in learning new words, discussions of the role of intentional (attentional) versus incidental (subconscious) acquisition processes, debates on the virtues and vices of teaching overt word learning strategies, and determinations of what makes one word more difficult to learn than another during reading. However, there have been relatively few investigations that have focused on the *actual vocabulary* involved in *actual extensive reading scenarios* to determine if the broad claims of the incidental hypothesis are linguistically viable—that is, to determine the extent to which typical extensive reading materials actually produce *multiple opportunities* for readers to encounter *new words* in *multiple contexts*. While it may be possible to discuss other important word-learning issues surrounding these three factors (i.e. contextual saliency, children's use of context cues, attention vs. no attention,

strategies vs. no strategies, and conceptual difficulty of words), it would be a mistake to lose sight of their central role in the claims being made, and an even more serious error to ignore them completely. In this regard, the current research may offer some unique perspectives on incidental vocabulary acquisition through extensive reading.

## THE STUDY

This study involves a lexical analysis of a corpus of fifth grade reading materials that 10- and 11-year-old children might encounter in their actual classrooms (Gardner 1999). The focus of the study is on findings relative to the narrative/expository variable in the corpus, with a secondary focus on thematic relationships between the texts in the corpus (for a more detailed treatment of thematic issues see Gardner 1999). The guiding questions for the remainder of the study are as follows:

1  To what extent do authentic narrative fiction and grade-equivalent expository reading materials share similar words?[8]
2  To what extent do authentic narrative fiction and grade-equivalent expository reading materials share similar word types at repetition levels predicted by theories of incidental vocabulary acquisition through extensive reading?

The findings will be used to examine the default claims of the *Incidental Acquisition Hypothesis*, as well as the pedagogical extensions of those claims—*Wide Reading* and *Free Reading*.

## The Corpus

The bulk of the corpus analysed in this study consists of texts from three popular thematic units used in US elementary education: *Mystery*, *Mummy*, and *Westward Movement* (see Appendix A). Of the 48 texts, 27 are from published thematic units. The expertise of a fifth grade teacher and a children's librarian were employed in choosing the 21 additional texts (not in published thematic units). Selections included subjective grade-level readability assessments (based on good to excellent readers), readability scores printed on the back covers of several books, popularity of texts, and thematic fit.

   The choice to use theme-related materials, some of which are written by a single author, is based on the assumption by many experts that such materials maximize the potential for words to be repeated (e.g. Brinton, Snow, and Wesche 1989; Chamot and O'Malley 1987: Enright and McCloskey 1988; Krashen 1985; Hwang and Nation 1989; Paribakht and Wesche 1999; Schmitt and Carter 2000; Stoller and Grabe 1997; Sutarsyah, Nation, and Kennedy 1994; Walmsley 1994)—a claim which itself may deserve more extensive

investigation (see Gardner 1999, for a more comprehensive treatment of this topic).

In addition to these thematic materials, collections of four thematically unrelated narratives (all popular *Newbery Medal* winners) representing different fictional registers (Science, Mystery, Adventure, and Romance) and four thematically unrelated expository texts representing four different content-areas (Earth Science, Political Science, Life Science, and Geography/ Culture) were also established with the aid of the two practitioners (see Appendix A for actual titles). The inclusion of these two non-thematic control collections raised the total number of children's texts in the corpus to 56 (28 narrative and 28 expository). All 56 texts were scanned into the computer using 'Omnipage' text scanning software and then carefully edited.

## Instruments for analysing the lexical compositions of text collections

The computer software deemed most suitable for addressing the research questions in this study were the 'VocabProfile' (VP) and 'Range' programs developed by Heatley, Hwang, and Nation of Victoria University's English Language Institute (see Laufer and Nation 1995, for further discussion). The three base lists accompanying these programs consist of the first and second 1,000 word families of the General Service List (GSL—West 1953) and approximately 835 word families of the University Word List (UWL). Both the VP and RANGE programs also allow researchers to create and use their own base lists instead of those accompanying the programs.

## Development of base lists for current study

Two base lists of higher frequency words were created for use with the VP and Range programs and successfully pilot tested (Gardner 1997). The goal in establishing these lists was to account for the most frequent words across a range of important corpora of the English language.

The source for Base List 1 is the study by Nation and Hwang (1995)[9] in which the researchers compared the highest frequency words of the GSL, Lancaster-Oslo Bergen (LOB) corpus of British English (Johansson 1978), and the Brown corpus (BRO) of American English (Francis and Kucera 1982). A computer diskette containing the overlapping and corpus-specific words identified in the 1995 study was obtained from Paul Nation. High frequency words appearing in any or all of the GSL, LOB, and BRO corpora *that were not also in the UWL corpus* were used to form Base List 1.

Base List 2 is comprised of the 835 word families from the UWL. To reiterate, the UWL is a distinct list from the GSL, and it has been found to have high frequency and range in adult-level academic texts (Xue and Nation 1984). It is included in this analysis to determine the impact of such academic words in reading materials targeted for 10- and 11-year-old children. While

no words in the UWL appear in the GSL, there are several that are found in the LOB and BRO high frequency lists. These were left in Base List 2 (UWL) and not included in Base List 1.

The two base lists were subsequently run against the 2,000 highest frequency word types of the American Heritage Intermediate (AHI)[10] corpus, and were shown to cover all but 94 of those 2,000 types (95.3 per cent), 70 of which were proper nouns (98.8 per cent coverage excluding proper nouns). The 94 types (along with their possible family members) were subsequently added to Base List 1 to strengthen the claim that the base lists represent the highest frequency words across many important corpora of the English language (see Table 2 for final breakdown of the base lists).[11]

*Table 2: Number of head words (families) and total word types by base list*

|  | Head words (Families) | Total Word types |
| --- | --- | --- |
| Base List 1 (GHF) | 2,350 | 8,510 |
| Base List 2 (AHF) | 835 | 3,672 |
| Total | 3,185 | 12,182 |

## Procedures for collecting lexical data

The computer programs enabled the researcher to characterize the reading materials based on three levels of vocabulary:

1 *General High Frequency Words (GHF)*—types from Base List 1 that represent the most frequent words of the English language across a range of subject areas and skill levels (e.g. 'the', 'of', 'and', 'bread', 'consider', 'example'). These are words that are expected to appear with *high frequency* in *many texts*.
2 *Academic High Frequency Words (AHF)*—types from Base List 2 (UWL) that occur with high frequency across a wide range of adult-level *academic* texts, but with less frequency than GHF words in general. Because AHF words are usually more specialized than GHF words, they are often referred to as 'subtechnical' words (e.g. 'abandon', 'absorb', 'academic').
3 *Unique Words (UNQ)*—types not found in either the GHF or AHF lists. Such types are often referred to as *technical* vocabulary if they appear frequently in a particular theme/subject area, or as *low frequency* if they do not (see Table 1). They also tend to be more cognitively challenging (e.g. 'kidnap', 'heist', 'crypt', 'archeologist', 'bison,' 'wrangler'). However, this level could also include more basic words such as names of characters in a series of books, certain idiosyncratic expressions used by those characters, names of famous people and places, and so forth.

It is important to note that GHF and AHF represent closed sets of words, whereas UNQ represents an open set that may vary from collection to collection, or text to text. Also, the pool of words from which UNQ types are drawn represents the bulk of the English word stock (see numbers of technical and low frequency words in Table 1). As such, they also represent both the challenge and the possibility of protracted vocabulary growth during the school years and beyond.

The reason for choosing distinct word types rather than word families or lemmas for analysis purposes is twofold: (1) while there is some research justification for assuming that young readers may be able to make word family connections, there is also ample evidence to suggest that this ability can vary from individual to individual depending on learner variables (L2 vs. L1, general reading skills, size of existing orally-based vocabulary, etc.), as well as the instructional interventions ('word consciousness' raising) they have been exposed to (Carlisle 2000; Freyd and Baron 1982; Mahony, Singson, and Mann 2000; Nagy, Diakidoy, and Anderson 1993; Leong 2000; Tyler and Nagy 1989); and (2) establishing word families among the large number of UNQ types in a corpus of this size would be a time-consuming and problematic undertaking.

## Repetition criteria

In addition to general type and token analyses, two word repetition criteria were used in comparing the lexical compositions of the narrative and expository registers:

1  *Zero Shared Frequency*—establishes the degree to which the narratives and grade-equivalent expository materials have exclusive vocabularies by determining how many words at the GHF, AHF, and UNQ levels appear (a) somewhere in the narratives, but *never* in the expositories, or (b) somewhere in the expositories, but *never* in the narratives. Considering words with zero shared frequency between narratives and expositories provides a solid control for those cases when only one or two key exposures to a new word might be enough for incidental vocabulary acquisition to take place (Landauer and Dumais 1997; Meara 1997). Results of this analysis will provide additional answers to the first research question of the study.
2  *10+ Shared Frequency*—establishes the degree to which words that *are* shared between narrative and expository materials actually reach facilitative levels of repetition (10+) for 'incidental' vocabulary acquisition to take place.[12] Results of this analysis will provide additional answers to the second research question of the study.

## RESULTS

Table 3 provides the word distribution data for the entire corpus. The 87.7 per cent token coverage for General High Frequency words (GHF) is in line with the 87 per cent predicted by Nation (1990) for such words. However, the 1.3 per cent token coverage for Academic High Frequency words (AHF) appears to be lower than would be expected, particularly given the fact that words at this level often account for 8 per cent of the total tokens in other reading materials. It is crucial to reiterate, however, that the AHF list has been shown to be highly salient in *adult academic* materials, not *adult fictional* materials (Nation 2001; Nation and Waring 1997). The total corpus figures in Table 3 do not take this distinction into consideration—a point that will be expanded upon shortly.

*Table 3: Token and type coverage in total corpus at three levels of vocabulary distribution*

|  | Tokens | % | Types | % | Families |
|---|---|---|---|---|---|
| General High Frequency | 1,265,780 | 87.7 | 7,167 | 21.8 | 2,322 |
| Academic High Frequency | 18,379 | 1.3 | 1,889 | 5.7 | 719 |
| Unique | 159,177 | 11.0 | 23,857 | 72.5 | |
| Total | 1,443,336 | 100 | 32,913 | 100 | |
| General High Frequency | 87.0 | | | | |
| Academic High Frequency | 8.0 | | Nation's (1990) Token | | |
| Unique | 5.0 | | Coverage Estimates | | |
| Total | 100 | | | | |

The combined GHF and AHF token coverage totals 89 per cent, with 11 per cent of all the tokens in the corpus being Unique (UNQ), meaning that they come from the large pool of lower frequency words that make up the bulk of the English word stock. Interestingly, however, the 159,177 occurrences of UNQ tokens (11 per cent of total) accounts for a staggering 72.5 per cent of the different types in the corpus (23,857 of 32,913 total), suggesting that UNQ types offer both the possibility and challenge for large scale vocabulary growth in these extensive reading materials—namely, many different word types to learn, but few chances to encounter those words in different contexts.

The figures in Table 4 shed additional light on the lexical distribution of word tokens in the corpus by comparing the 28 narrative texts with the 28 expository texts. It is clear from this comparison that the narrative and expository registers utilize vocabulary in very different ways. For instance, children's narratives in this corpus tend to contain a higher percentage of GHF

words than children's expositories (88.7 per cent to 84.9 per cent), whereas expositories tend to contain higher percentages of AHF words than narratives (2.7 per cent to 0.8 per cent) and higher percentages of UNQ words than narratives (12.4 per cent to 10.5 per cent). Because lexical specialization increases from left to right in the table, these figures suggest that expositories create more difficult reading experiences for young readers, and that word learning in expository reading would be more difficult in general (i.e. more knowledge of specialized words needed to unpack the meanings of other specialized words).

The disparities at the AHF level also help to clarify the low AHF coverage percentage noted in Table 4. While the 2.7 per cent AHF coverage for children's expository materials is far below Nation's 8 per cent figure for adult academic materials, it is still much higher than the 0.8 per cent for children's narratives. This suggests that text type (narration vs. exposition) begins to affect children's exposure to academic vocabulary as early as the fifth grade (and possibly earlier).

In order to determine if the large disparities in numbers of total tokens between the 28 narratives (1,057,976 tokens) and 28 expositories (385,360 tokens) has an effect on the lexical distributions, it was also necessary to analyse equal sample sizes of narrative and expository material. Table 5 reports the lexical coverage figures when only the first 5,000 tokens of each book are considered (140,000 tokens of narrative material and 140,000 tokens of expository material). A comparison of Tables 4 and 5 clearly demonstrates that the lexical distributions of children's narrative and expository materials are very consistent, regardless of sample size, and that the lexical differences noted between the two broad registers are valid—namely, that children's narratives draw more heavily on GHF words and less heavily on AHF and UNQ words than children's expositories in carrying out their discourse functions.

Obscured in the general token data, however, are the lexical characteristics relative to individual word types. That is, the token coverages do not give specific information regarding either actual or average repetitions of individual words types within the various levels of distribution.

*Table 4: Average token coverage in 28 narrative and 28 expository texts at three levels of vocabulary distribution (using all tokens in corpus)*

| | % GHF Tokens | % AHF Tokens | % UNQ Tokens |
|---|---|---|---|
| NARRATIVE (28 Texts) 1,057,976 Tot Tokens | 88.7 | 0.8 | 10.5 |
| EXPOSITORY (28 Texts) 385,360 Tot Tokens | 84.9 | 2.7 | 12.4 |

*Table 5: Average token coverage in 28 narrative and 28 expository texts at three levels of vocabulary distribution (using first 5000 tokens per text)*

|  | Avg % GHF Tokens | Avg % AHF Tokens | Avg % UNQ Tokens |
|---|---|---|---|
| NARRATIVE (28 Texts) 140,000 Tot Tokens | 88.5 | 0.7 | 10.8 |
| EXPOSITORY (28 Texts) 140,000 Tot Tokens | 85.1 | 2.7 | 12.2 |

As the figures in Table 6 suggest, there are also marked differences between the children's narrative and expository reading materials in terms of the number of *different* word types that children will encounter. In these sets of extensive reading materials, a child who reads 5,000 tokens (approximately 20 pages) of expository material will be required to negotiate 135 more word types, on average, than a child who reads 5,000 tokens of grade-equivalent narrative fiction. Interestingly, this lexical burden in expository reading occurs at all three levels of vocabulary distribution, even at the GHF level (55 more GHF types on average in expositories than in narratives), despite the fact that narratives actually have higher token percentages than expositories at the GHF level (Narratives, 88.5 per cent; Expositories, 85.1 per cent—see Table 5). These lexical figures clearly support the on-going conclusion that expository reading materials pose different lexical challenges in the context of extensive reading, and that they are much less likely than narratives to offer the conditions of 'easy', 'pleasureful', and 'i + 1' associated with incidental word acquisition through reading.

*Table 6: Average number of different word types in the first 5,000 tokens of 28 narrative and 28 expository texts at three levels of vocabulary distribution*

|  | #GHF | #AHF | #UNQ | Total |
|---|---|---|---|---|
| Narrative | 501 | 24 | 239 | 764 |
| Expository | 556 | 61 | 282 | 899 |
| Net Difference | 55 +exp | 37 +exp | 43 +exp | 135 +exp |

## Zero frequency-type non-sharing between narrative and expository texts in total corpus

While the vocabulary findings discussed thus far seem to point to children's narratives as providing more optimal lexical conditions for incidental word

*Table 7: Number of types never shared (zero frequency) between narrative & expository registers at three levels of vocabulary*

| Vocab Level | # Types Corpus | # of Types exclusive to narratives (0 Freq in Exp) | % of total | # of Types exclusive to expositories (0 Freq in Nar) | % of total | # of Types exclusive to Nar or Exp | % of total |
|---|---|---|---|---|---|---|---|
| GHF | 7,167 | 759 | 10.6% | 691 | 9.6% | 1,450 | 20.2% |
| AHF | 1,889 | 260 | 13.8% | 563 | 29.8% | 823 | 43.6% |
| UNQ | 23,857 | 10,041 | 42.1% | 7,250 | 30.4% | 17,291 | 72.5% |
| Total | 32,913 | 11,060 | 33.6% | 8,504 | 25.8% | 19,564 | 59.4% |

acquisition than children's expositories, the first major question of the current study nevertheless remains unaddressed—namely, to what extent are the words in the easier narratives *the same* as those in the more lexically-dense expository materials?

The figures in Table 7 provide important insights regarding this issue. To begin, it is clear that the total percent of non-sharing between the 28 narratives and 28 expository texts increases as vocabulary levels become more specialized.

(GHF—20.2 per cent to AHF—43.6 per cent to UNQ—72.5 per cent). Even at the General High Frequency level (GHF), where types are assumed to have frequency and range across texts in general, 1,450 types (20.2 per cent) are register specific—that is, they do not occur at all in 28 narratives, or they do not occur at all in 28 expository texts. Nearly half (43.6 per cent) of the Academic High Frequency types (AHF) and approximately three-quarters (72.5 per cent) of the words not in the high frequency or academic lists occur in one macro-register only. In fact, of the 23,857 UNQ types in this corpus of 1,443,336 total tokens, only 6,566 have any register cross-over whatsoever, while 17,291 are specific to either narrative or expository texts.

Perhaps even more telling than the staggering numbers of non-shared types are the differences in characteristics of the types themselves. Table 8 contains examples of the actual types in the corpus that are exclusively narrative or exclusively expository. A visual inspection of these words makes it clear that the narrative types tend to emphasize human characteristics (e.g. emotions, expressions, physical features), and the social and cultural contexts of human interaction. In stark contrast, the expository types are more informational, scientific, and content-based in nature. Examples of General High Frequency types (GHF) that are specific to the broad narrative register include *frighten, grasped, hi, breathless, creep, fooling, hating, interrupting, patiently, pleasantly*, and *screwed*, while examples characteristic of grade-equivalent expository types include *continues, diseases, experiments, measurements, engineers, gases, minerals, advances, discoveries, governments, markets*, and *organization*.

Likewise, at the other extreme vocabulary level (UNQ), examples characteristic of narrative types include *blinked, dangled, growled, hug, huh, sleeve, whined, cab, flopped, giggled, jerking*, and *puff*, while examples characteristic of grade-equivalent expository types include *expeditions, exposure, irrigation, abundance, artisans, colossal, conquest, eruption, excavated, fahrenheit, glossary*, and *inhabited*.

These characteristic differences, together with the raw counts of non-shared types, strongly support the ongoing conclusion of this study—namely, that the lexical compositions of children's narrative and grade-equivalent expository reading materials are largely dissimilar, particularly at the more specialized levels of vocabulary (UNQ) where the potential for large-scale vocabulary acquisition is the greatest.

*Table 8: Examples of register specific types by vocabulary level, register, collection range, and alpha listing within range*

| GHF | | AHF | | UNQ | |
|---|---|---|---|---|---|
| Narrative Only | Expository Only | Narrative Only | Expository Only | Narrative Only | Expository Only |
| (6)* | (6) | (5) | (6) | (7) | (7) |
| CUSTOMER | CONTINUES | COINCIDENCE | CONSTRUCTED | ATTIC | EXPEDITIONS |
| DROPPING | DISEASES | LIABLE | (5) | BLINKED | (6) |
| FRIGHTEN | EXPERIMENTS | SWITCHED | AFFECTED | BRACED | AUSTRALIA |
| GRASPED | MEASUREMENTS | (4) | GUARANTEED | CLAMPED | EUROPEANS |
| HI | MEDICINES | COMMENT | ILLUSTRATION | CUFFS | (5) |
| INTERRUPT | MINERAL | CONCENTRATION | ISSUES | DANGLED | ALFRED |
| SHUTTING | SHORTAGE | REACTION | REIGN | FLAPPED | ANTARCTIC |
| SMILES | (5) | RESPONDED | (4) | GLARED | BEST-KNOWN |
| SNAPPING | CANADIAN | SHIFT | ACHIEVE | GRABBING | BRITAIN |
| STROKING | CELEBRATED | TEMPORARILY | CHAPTERS | GROWLED | EMPEROR |
| SWEATING | DESERTS | TENSION | CONDUCTED | HISSED | EXPOSURE |
| TALKS | ENGINEERS | TINIEST | CRYSTALS | HUG | GULF |
| WEAKLY | GASES | (3) | CULTURES | HUH | IRRIGATION |
| WIPING | GOATS | ANGULAR | DECADE | KNOB | MEDITERRANEAN |
| YOURSELVES | MARCHES | APPRECIATION | EDITOR | LATCH | OKLAHOMA |
| (5) | MINERALS | ASSURANCE | ENABLES | LUNGED | SURVEYORS |
| ANNOUNCEMENT | MUSICIANS | AUTOMATIC | ENSURE | NODDING | (4) |
| APPOINTMENT | OFFERS | BENEFACTOR | EVOLVED | PEEKED | ABUNDANCE |
| BLESS | ORGANS | BIOLOGY | FRONTIER | SLEEVE | ADAMS |
| BREATHLESS | READS | COMMISSIONER | HERITAGE | SMARTER | ALLAN |
| CREEP | RICHARD | CONCENTRATING | INVESTIGATIONS | STACK | AMAZON |
| DANGEROUSLY | (4) | CONSULTING | MIGRATION | TRAILED | ARTISANS |
| DOORSTEP | ADDS | CONTRARY | NETWORK | WEARILY | BACKGROUNDS |
| DRIVEWAY | ADVANCES | DISTINCTLY | PORTS | WHINED | BANISHED |
| EVENLY | ATTRACTS | EMBRACE | RANGED | (6) | BANNED |
| FOOLING | AUDIENCES | EMPHASIS | REVOLUTIONARY | AISLE | BRAVERY |
| FRIED | BLACKS | EVIDENTLY | STATUS | BARKED | COLONEL |
| HATING | DECAY | INHERITANCE | SURVEYS | BOUNCING | COLOSSAL |
| HURTS | DESCRIBES | INITIALS | TECHNIQUE | CAB | CONQUEST |
| INDOORS | DIRECTORS | INSIST | TREATY | CLATTERED | DESTINED |
| INTERRUPTING | DISCOVERIES | INVOLVEMENT | VERSIONS | CLICK | ERUPTION |
| MOTIONS | EXTENSIVE | MOTIVE | (3) | CLICKING | EXCAVATED |
| MUSTN'T | GOVERNMENTS | PREMISES | ADAPTED | CLUMPS | FABULOUS |
| NEEDING | GUIDES | RECTANGLE | AFFECT | CREAK | FAHRENHEIT |
| PARDON | IMPROVEMENTS | SCORE | AFFECTS | FLOPPED | FESTIVAL |
| PASTED | INVENTIONS | TENSELY | AIDED | FROWNING | FESTIVALS |
| PATIENTLY | LEADERS | TONES | APPEALED | GIGGLED | FISHERMAN |
| PILING | MARKETS | TOPIC | ASSIST | GLARING | GANGS |
| PLEASANTLY | ORGANIZATION | VIBRATED | AUTHORS | HUGGING | GLOSSARY |
| PUMPING | PERFORMING | (2) | CAPTURES | JERKING | HARPOON |
| SCREENED | PHOTOGRAPHERS | ACCUMULATED | CHALLENGER | MOURNFUL | HEYDAY |
| SCREWED | PHOTOGRAPHY | BOMBER | CHALLENGING | OWL | INHABITED |

*Table 8 (cont.)*

| GHF | | AHF | | UNQ | |
|---|---|---|---|---|---|
| Narrative Only | Expository Only | Narrative Only | Expository Only | Narrative Only | Expository Only |
| (6)* | (6) | (5) | (6) | (7) | (7) |
| SLOPED | PIGEONS | CAREERS | COMMITMENT | PUDDLE | INTESTINES |
| SOCK | POPULARITY | CATERING | COMMUNITIES | PUFF | IRELAND |
| SPOIL | PRODUCES | CLIENT | CONGRESS | RACKET | JUNGLE |
| STATEMENT | PROFITS | COMPLICATIONS | CONSTITUTIONAL | RUSTLED | LEGENDARY |
| STOCKING | QUANTITIES | CONCLUDE | CONVENTION | SCORCHED | LIFE-SIZE |
| STUNG | QUEST | CONSEQUENCE | CYCLE | SCRAMBLING | LOUISIANA |
| THICKER | RELATION | CONSOLATION | DEMOCRACY | SHOVING | MAMMALS |
| TIDY | RESULTED | CONTRACTED | DEMOCRAT | SKIPPING | MEMPHIS |
| TRIPPED | SEAMEN | CONTRADICTORY | DEMOCRATIC | SNORING | MI |
| | SLAVERY | CRITICIZING | DISPUTES | | MIDDAY |
| | | DEBATED | DISTINCTIVE | | MO |

\* Number of collections in register containing listed types. each collection consists of 4 books. max. collections per register = 7

## Shared types between narratives and expositories with 10+ repetitions

Despite the noted lexical differences between narratives and expositories in the corpus, it could still be argued that considerable sharing does take place, as indicated in Table 9 (i.e. 5,717 shared GHF types, 1,066 shared AHF types, 6,566 shared UNQ types). However, beyond the General High Frequency level (GHF), only 311 shared types (78 AHF + 233 UNQ) reach facilitative incidental word acquisition levels (10+ repetitions) in both the narrative and expository registers.

*Table 9: Number of shared types between narrative and expository registers with 10+ repetitions in each register*

| General high frequency | | | Academic high frequency | | | Unique | | |
|---|---|---|---|---|---|---|---|---|
| Total Shared | 10+ Shared | % of Total shared | Total Shared | 10+ Shared | % of Total shared | Total Shared | 10+ Shared | % of Total shared |
| 5,717* | 2,374 | 41.5% | 1,066* | 78 | 7.3% | 6,566* | 233 | 3.5% |

And without thematic relationships between the narrative and expository materials (i.e. *Mystery*, *Mummy*, *Westward Movement*), the actual number of shared UNQ types meeting the 10+ repetition criterion would be very low indeed. For instance, a visual inspection of these shared types (see Table 10) makes it clear that the *Mummy* theme accounts for the words *Cairo*, *Canopic*, *Egypt*, *Egyptian*, *Egyptians*, *Isis*, *mummies*, *mummy*, *pharaoh*, *priestess*, *pyramid*, *pyramids*, *sarcophagus*, *scarab*, *tomb*, *tombs*, and *wrappings*. *Mummy* or the looser *Mystery* theme might also account for the words *artifacts*, *bandages*, *beads*, *belly*, *bronze*, *candles*, *carved*, *cemetery*, *coffin*, *corpse*, *decorated*, *diggings*, *eternity*, *ghost*, *ghosts*, *granite*, *inscription*, *invisible*, *jewelry*, etc. And *Westward Movement* likely accounts for the words *beef*, *buffalo*, *bullets*, *calf*, *calves*, *canoe*, *cowboys*, *creek*,

*Table 10: 233 UNQ types with 10+ repetitions in both narrative and expository registers*

| | | | | | | |
|---|---|---|---|---|---|---|
| ABOARD | CHASE | FIREPLACE | JEWELRY | NILE | RELIEF | SURVIVE |
| AD | CHICAGO | FLAIL | JOHNSON | NODDED | RIBS | SUSAN |
| ALEXANDER | CHIN | FLORIDA | KA | NOTED | RIFLE | TERRIFIED |
| ANNA | CHIPS | FORD | KANSAS | OAK | ROBBERS | TERRIFYING |
| ARTHUR | CHUNKS | FOREHEAD | LA | OREGON | ROBES | TH |
| ARTIFACTS | CIRCUS | FRANCISCO | LAB | OUTFIT | ROCKY | THEATER |
| ARTS | COALS | FROZE | LEWIS | OX | ROSS | THOMPSON |
| BANDAGES | COFFIN | GEAR | LINED | OXEN | SAMUEL | THOTH |
| BEADS | CONNECTICUT | GHOST | LINEN | PALM | SAN | THRONE |
| BEEF | CORPSE | GHOSTS | LOUIS | PANTS | SANTA | TIGHTLY |
| BELLY | COWBOYS | GLOWING | LUCY | PASTURE | SARCOPHAGUS | TIMBER |
| BELONGINGS | CRADLE | GOLD-SEEKERS | LUKE | PATROL | SCALP | TOMB |
| BENTON | CREEK | GRABBED | LUMBER | PEAK | SCARAB | TOMBS |
| BLAST | CROOK | GRANITE | MANSION | PHARAOH | SCARED | TOSSED |
| BLUFFS | CROW | GRAVEL | MARBLE | PIT | SCARY | TRADER |
| BOOM | DAKOTA | GRIZZLY | MARSHALL | PLANKS | SCOTT | TROUSERS |
| BOOTS | DAVID | HANDY | MASK | PLASTIC | SCOUT | TRUDGED |
| BRONZE | DEADLY | HAULED | MASSIVE | PLATFORM | SEALED | TUNNEL |
| BUFFALO | DECORATED | HAULING | MICHIGAN | PLOW | SEARCHERS | TURTLE |
| BULL | DENTIST | HAUNTED | MIGHTY | PONY | SHAFT | TV |
| BULLETS | DIARY | HAWK | MINNESOTA | PORK | SHOVELS | TWENTY-FOUR |
| BUN | DIGGINGS | HERBS | MISSISSIPPI | PORTRAIT | SHROUD | UNWRAPPED |
| BURT | DOLL | HERD | MISSOURI | PRAIRIE | SILLY | VAN |
| CABINET | DOLPHINS | HOMESTEAD | MONSTER | PRICKLY | SIOUX | WEBSTER |
| CAIRO | DUSTY | HOWARD | MOSS | PRIESTESS | SKULL | WEIRD |
| CALF | EGYPT | INSCRIPTION | MOUND | PRINCE | SLAB | WILDERNESS |
| CALVES | EGYPTIAN | INVISIBLE | MULE | PROFESSOR | SLIT | WILSON |
| CANDLES | EGYPTIANS | IOWA | MUMMIES | PYRAMID | SOAKED | WORKMEN |
| CANOE | ELEGANT | ISIS | MUMMY | PYRAMIDS | SOD | WOVEN |
| CANOPIC | ELIZABETH | IVORY | MUSEUM | QUARTZ | STATUE | WRAPPINGS |
| CARGO | EMERSON | JACKSON | NAKED | QUIT | STATUES | YANKEE |
| CARVED | ETERNITY | JAIL | NEBRASKA | RAMP | STICKY | YOKE |
| CEMENT | FIN | JESSE | NECKLACE | RATS | STRANDED | YORK |
| CEMETERY | FINGERPRINTS | | | | | |

## Table 11: Highest frequency UNQ types by register and frequency

| 1st 50 Narrative | | 2nd 50 Narrative | | 1st 50 Expository | | 2nd 50 Expository | |
|---|---|---|---|---|---|---|---|
| LAURA | 1988 | MICHAEL | 204 | MONTANA | 565 | YELLOWSTONE | 63 |
| PA | 1893 | GRAMPA | 199 | **MUMMY** | 479 | JOSEPH | 62 |
| MA | 1526 | BERRYMAN | 196 | MUMMIES | 393 | PHARAOH | 62 |
| ROSIE | 1443 | BERT | 185 | **EGYPT** | 328 | LINEN | 61 |
| KAYO | 1386 | MUFFIN | 179 | PYRAMID | 233 | GHOST | 59 |
| ANTHONY | 1316 | THERESA | 178 | TOMB | 213 | PEAT | 58 |
| EELLS | 973 | STATUE | 177 | EGYPTIANS | 211 | HELENA | 57 |
| ALMANZO | 888 | GRABBED | 176 | MISSOURI | 206 | LOUIS | 56 |
| PROFESSOR | 860 | KNIEVEL | 175 | SKULL | 176 | VIRGINIA | 56 |
| PHINEAS | 736 | ALICE | 174 | LUKE | 173 | LEWIS | 55 |
| MEG | 638 | FLASHLIGHT | 171 | EGYPTIAN | 160 | MISSISSIPPI | 55 |
| TONY | 560 | YEAH | 170 | PYRAMIDS | 157 | EMBALMING | 54 |
| PRAISEWORTHY | 480 | CAROLINE | 168 | SAN | 156 | PANAMA | 53 |
| ALTHEA | 439 | GRAMMA | 168 | TOMBS | 147 | PP | 53 |
| CLAUDIA | 431 | SODDY | 163 | YORK | 138 | FE | 52 |
| MOMMA | 420 | BORKMAN | 162 | FRANKLIN | 130 | SEALED | 52 |
| ADDIE | 409 | FERGIE | 159 | OREGON | 121 | COFFINS | 51 |
| JAMIE | 404 | POPPA | 159 | TH (AS IN 20TH) | 118 | MASK | 51 |
| WESTING | 362 | ELENA | 156 | SHUTTLE | 114 | NEVADA | 51 |
| **PRAIRIE** | 355 | WEXLER | 156 | FRANCISCO | 110 | ROCKY | 51 |
| TAD | 335 | SIGHED | 155 | PETRONIA | 110 | CHARNEY | 50 |
| WALLACE | 335 | YELLED | 155 | COFFIN | 109 | DYNASTY | 50 |
| CALVIN | 334 | CROW | 151 | COWBOY | 105 | BUTTE | 49 |
| SUSANNAH | 331 | LETTY | 150 | ARCTIC | 93 | CARVED | 49 |
| MELANIE | 328 | ANNA | 141 | BUFFALO | 91 | ASTRONAUTS | 48 |
| LESLIE | 322 | ELIZA | 139 | **MUSEUM** | 87 | DECORATED | 47 |
| CREEK | 311 | HUGO | 138 | EXPEDITION | 85 | TRAVELERS | 47 |
| TURTLE | 311 | SAUNDERS | 138 | SPERM | 82 | AUDUBON | 46 |
| JESS | 285 | VAN | 138 | FORENSIC | 81 | COLORADO | 46 |
| LESTER | 281 | EMERSON | 137 | BOG | 80 | COLUMBIA | 46 |
| KEN | 276 | SYDELLE | 135 | CLARK | 80 | OHIO | 46 |
| ALEXANDER | 275 | THEO | 133 | NILE | 79 | PHARAOHS | 46 |
| DAVID | 269 | COOKSON | 131 | BALEEN | 77 | **PRAIRIE** | 46 |
| WHATSIT | 265 | GRINNED | 130 | WHALING | 77 | BILLINGS | 45 |
| NODDED | 264 | SAMMY | 128 | OSIRIS | 76 | EXPLORERS | 45 |
| CARRIE | 263 | BET | 127 | BURIAL | 75 | JACKSON | 45 |
| **MUMMY** | 259 | DOC | 123 | COWBOYS | 73 | RESERVATION | 45 |
| SHEBA | 258 | PORCH | 123 | INVISIBLE | 71 | ROBBERS | 45 |
| **EGYPT** | 248 | AMBER | 120 | MUMMIFIED | 71 | STATUES | 45 |
| SCARED | 246 | PEERED | 120 | SANTA | 71 | THOMAS | 45 |
| OKAY | 242 | OTIS | 118 | EMIGRANTS | 68 | TRADERS | 45 |
| TOBY | 242 | WEBSTER | 117 | KANSAS | 68 | BLUBBER | 44 |
| CHRIS | 227 | FAIRWEATHER | 116 | PERCENT | 67 | DIGGINGS | 44 |
| ANGELA | 223 | MURRY | 116 | BC | 65 | HERD | 44 |
| LYLE | 221 | MARTHA | 115 | HERCULANEUM | 65 | IOWA | 44 |
| QUIGGY | 221 | MUTTERED | 115 | OVERLAND | 64 | MULES | 44 |
| **MUSEUM** | 216 | BELLE | 113 | SMITH | 64 | UTAH | 44 |
| MARSHALL | 214 | BOOTS | 113 | FINGERPRINTS | 63 | JEKYLL | 43 |
| TALLIE | 213 | FORD | 113 | OXEN | 63 | VESUVIUS | 43 |
| HOO | 205 | AIN'T | 112 | SUTTER | 63 | HORUS | 42 |

*Dakota, herd, homestead, jail, Kansas, ox, pasture, plow, prairie, Sioux, trousers, wilderness,* etc.[13]

Perhaps Table 11 provides the clearest illustration of the problematic nature of word sharing between children's narrative and expository materials. Upon visual inspection, it becomes apparent that the highest frequency UNQ types are vastly different in narratives and expositories. In narratives, these types tend to be *characters* (*Laura, Pa, Ma, Rosie, Kayo, Anthony, Eells, Almonzo,* etc.),[14] *places* (*prairie, creek, museum, porch,* etc.), *simple action verbs* (*nodded, grabbed, yelled, grinned, peered,* etc.) or *expressions* (*okay, yeah,* etc.).

In short, they are the words of story-telling. In expositories, on the other hand, the UNQ types tend to be much more content-specific and theme-related in nature—mostly nouns (*Montana, mummy, mummies, Egypt, pyramid, tomb, cowboy, buffalo, expedition, bog, cowboys, oxen, pharaoh, dynasty, diggings,* etc.). Furthermore, only four UNQ types in the top 100 of both registers are actually shared (*mummy, Egypt, museum,* and *prairie*)—all four being heavily tied to themes.

Two additional differences between the narrative and expository types seem to stand out as well:

1   None of the top 100 narrative types appear to be from the same morphemic families, with the possible exception of certain relationship terms (*Pa, Poppa, Grampa;* and *Ma, Momma, Gramma*). In contrast, of the top 100 expository types, 16 have distinct family relationships (*mummy, mummies, mummified; Egypt, Egyptians, Egyptian; pyramid, pyramids; tomb, tombs; coffin, coffins; cowboy, cowboys;* and *pharaoh, pharaohs*).
2   Virtually none of the top 100 narrative types represent complex concepts, but many of the top 100 expository types do (*expedition, forensic, whaling, mummified, percent, embalming, dynasty, diggings,* etc.).

Finally, it should be noted that even when certain UNQ types are shared between children's narratives and expositories, they may not receive the same emphasis, or even have the same meaning, in both macro-registers. For instance, the word *mummy* appears to be an important shared type in the corpus under investigation. However, examples of actual contexts containing the word *mummy* clearly demonstrate a narrative-expository distinction in terms of the way *mummy* is employed in the two broad registers (see Table 12). In the narrative examples, it appears that *mummy* merely contributes to the larger story-telling functions of a narrative—namely, the development of plot, setting, characterization, etc. In a sense, the word *mummy* in such materials becomes incorporated, rather than highlighted, in the overall narrative structure. And the semantic notion of *mummy* in the narrative contexts appears to be a Hollywood-like mixture of fiction and fact.

In contrast, the grade-equivalent expository contexts appear to treat *mummy* in the sense of terminology, explicating the same through definition and explanation (likely in conjunction with diagrams, pictures, glossaries, etc.). In other words, the *mummy term* is actually the reason why the expository

*Table 12: The word 'Mummy' in narrative and expository texts*

| Narratives | Expositories |
|---|---|
| **The Mummy, the Will, and the Crypt (Bellairs, 1983)** | **Mummies & Their Mysteries (Wilcox, 1993)** |
| 'The young man paused and grinned unpleasantly. 'Do you know what a *mummy* looks like after it's been unwrapped? Just a dried brown husk that used to be a human being, with holes for eyes?'' [emphasis added] | 'A *mummy* is the body of a human or animal in which some of the soft tissues (skin, muscles, or organs) did not decay after death. This makes a *mummy* different from a skeleton or a fossil.' [emphases added] |
| 'The other was stretched out, and his hand was layed flat on the floor. It was brown and withered, like the hand of a *mummy*.' [emphasis added] | 'Drying isn't the only way to turn a body into a *mummy*. Taking away all air from around the body will stop decay, since bacteria and fungi need air as well as water to live.' [emphasis added] |
| 'A figure with hollow *mummy* eyes and a withered *mummy* face and clawlike *mummy* hands. Moving with an awful, tottering, unsteady gait, it came toward him.' [emphases added] | 'When the word *mummy* was first used in the English language in the early 1400s, it did not mean a body as it does now. Instead, it was the name of a medicine. *Mummy* comes from mumiyah, an Arabic word for bitumen, a sticky oil now used to make roads.' [emphasis added] |
| **The Vandemark Mummy (Voight, 1992)** | |
| 'Phineas looked across the table, across the *mummy*'s face, to his father. Mr. Hall stared down at the wrapped figure, and at the portrait face that was held in place by wrappings, and then back down the length of the *mummy* to its feet, where Althea stood staring.' [emphases added] | **Tales Mummies Tell (Lauber, 1985)** |
| | 'The ground froze and stayed frozen, except for the top few inches, which thawed each summer. In this natural deep freeze, the body of the baby mammoth was preserved for thousands of years. It became a *mummy*, which is the term now used for any well-preserved body, whether animal or human.' [emphasis added] |
| 'He wasn't sure he exactly understood that now, but he felt as if this ancient *mummy* stood for something truer than . . . all the money he could imagine winning in the lottery, truer than Donald Trump, truer even than the threat of nuclear war and nuclear accidents, AIDS, or the waste crisis.' [emphases added] | 'To find the age of the *mummy*, scientists made use of a built-in atomic clock. This is how the clock works: Certain kinds of atoms are radioactive they keep breaking down by giving off tiny parts of themselves. Among these atoms are those of carbon 14, which is a radioactive variety of carbon.' [emphasis added] |
| '''The Collection, as you will find, is a hodgepodge. There will be some pleasant surprises for you, or so I like to think. The *mummy*, which is its centerpiece, has a certain wistful appeal, being from the Roman era.''' [emphasis added] | 'Copper Man is a South American *mummy* that formed naturally with the help of dry air and salts. This *mummy* was once a copper miner who lived and worked around A.D. 800 in the Atacama Desert of northern Chile.' [emphasis added] |

structures exist in the first place, and the expository structures, in turn, define or explain the *mummy* term. The same cannot be said of the relationship between *mummy* and the narrative structures in which it is embedded.

## DISCUSSION

The reading materials analysed in this study clearly contain large volumes of specialized English words. In fact, of the 32,913 total types in the corpus, 23,857 (72.5 per cent) are outside the high frequency lists. On the surface, these statistics provide corollary support to numerous studies which have linked children's extensive exposure to written material and growth in their vocabulary knowledge (e.g. Krashen 1989; Miller and Gildea 1987; Nagy and Anderson 1984; Nagy, Anderson, and Herman 1987; Nagy and Herman 1984, 1987; Nagy, Herman, and Anderson 1985; Stanovich *et al.* 1996). However, at a more precise level, the findings clearly indicate that *all reading is not the same*, and that the choice of texts which children are exposed to during extensive reading (e.g. narrative vs. expository; texts related by theme vs. texts not related by theme) will have a profound effect on the types of words they can potentially learn, the number of times they will encounter certain types, and the amount of prior vocabulary knowledge they will need to actually learn new words during extensive reading experiences. In short, the claim of large scale incidental vocabulary acquisition through extensive reading experiences ('sheer volume') forwarded by advocates of *Free Reading* and *Wide Reading* appears to be much too broad to account for these factors. Several key findings in this study serve to illustrate this assertion.

### Findings that support the *Incidental Acquisition Hypothesis*

1. Children's narratives tend to utilize a greater proportion of General High Frequency (GHF) words than children's expositories, while expositories tend to utilize a greater proportion of words at the more specialized vocabulary levels (i.e. Academic High Frequency (AHF) and Unique (UNQ) words).[15] This favours narrative contexts over expository contexts for incidental word acquisition because it indicates that narratives place fewer lexical demands on children in general.

2. Given equivalent amounts of reading material, children's narratives contain far fewer different word types than expositories (on average, 135 fewer types than children's expositories in 5,000 total words—approximately 20 pages of reading material). Furthermore, this disparity in different types over 5,000 tokens distributes itself across all levels of vocabulary, including the General High Frequency level (GHF = 55 more types in expositories; AHF = 37 more types in expositories; UNQ = 43 more types in expositories). Again, these findings favour narrative contexts over expository for incidental word acquisition because of reduced lexical demands placed on the reader.

These first two major findings provide some linguistic support for the 'good

stories' position taken by advocates of incidental vocabulary acquisition, and possibly suggest how the requisite conditions for such acquisition ('comprehensible input', 'i + 1', 'pleasureful reading', etc.) are more easily met when readers are involved in narrative reading experiences (Anderson 1996; Krashen 1989).

## Findings that do not support the *Incidental Acquisition Hypothesis*

1. The words available for potential acquisition during children's narrative reading are vastly different than those potentially acquired during children's expository reading. Of the 23,857 UNQ types in the corpus, 17,291 (72.5 per cent) are register specific—that is, they appear somewhere in the narratives, but *never* in the grade-equivalent expositories, or they appear somewhere in the expositories, but *never* in the grade-equivalent narratives. This finding becomes even more noteworthy when one considers that the potential for large scale vocabulary growth during the school years must surely involve words at this particular level (i.e. the bulk of the English word stock).[16] Also important is the fact that the disparities in UNQ vocabulary between narratives and expositories exist despite the fact that many of the books containing these words are supposedly related to each other by a common theme.

  Therefore, given the higher lexical density of most expository materials—a phenomenon not typically associated with 'i + 1' or free pleasure reading—how might the large numbers of specialized vocabularies in the expositories be acquired during the academic years? This is a question that seems to fuel many discussions among proponents of content-based instruction in L2 settings (e.g. Chamot and O'Malley 1987; Grabe and Stoller 1997; Kinsella 1997; Raptis 1997; Carrell and Carson 1997) and appears to inspire, in part, the research and curricular positions of many content-area reading specialists in L1 contexts (e.g. Deshler, Ellis, and Lenz 1996; Vacca and Vacca 1996).

  To reiterate, if so many types are *never shared* between narrative and expository texts in nearly 1.5 million tokens of grade-equivalent, theme-related, reading materials, it seems difficult to justify the claim that they will eventually do so at the times when the children need them the most (i.e. during grade-level reading), and even more difficult to justify that they will eventually do so at the levels of repetition that *Free Reading* advocates claim are necessary for 'incidental' learning to take place (e.g. 10+ repetitions). Furthermore, the findings relative to the unique words of difficult expositions (contexts which may be 'i + 5?' or 'i + 10?' rather than 'i + 1') certainly do not favour the assertion by many L2 extensive reading experts, including Krashen (1989) and Elley (1991), that children will have learned 'all the vocabulary and syntax they required in due course from repeated interactions with good stories' without 'regular analytic study and practice with the language' (Elley 1991: 378, 379).

  2. The register-specific UNQ types are characteristically different. The

narrative UNQ types tend to emphasize human characteristics (e.g. emotions, expressions, physical features), and the social and cultural contexts of human interaction. In stark contrast, their expository counterparts are more abstract, informational, and content-based in nature—in short, they are more characteristic of the language of school.

One might argue here that the exclusively expository types in this corpus may eventually appear in narratives read at higher grade levels, but the characteristic differences noted in the words themselves (i.e. narrative words tending to describe basic human characteristics, expressions, settings, conflicts; expository words tending to describe more abstract content concepts) suggest that the differentiation may become even greater, especially as content topics become more specialized in later grades.

3. The bulk of the UNQ types that are shared between the two broad registers (6,333 of 6,566—96.5 per cent) tend not to reach facilitative levels of repetition predicted by the *Incidental Acquisition Hypothesis*—that is, they do not have 10+ repetitions across the range of narrative texts and also 10+ repetitions across the range of expository texts. Again, this is a particularly sticky problem for the *Incidental Acquisition Hypothesis* because it suggests that many of the specialized words that are shared between narratives and grade-equivalent expositories only appear with sufficient repetition (10+) in nonfacilitative contexts ('more difficult expositions'), or they appear with sufficient repetition in facilitative contexts ('easy narratives'), but they do not play an important lexical role in the nonfacilitative contexts (i.e. they appear less than 10 times overall in 28 expository books).

Furthermore, these low shared repetition figures do not favour the *unqualified Wide Reading* claims of *multiple encounters* with the same words in *multiple contexts* through *sheer volume* of reading experiences, especially if a reasonable time-frame (Nation 1997) is taken into consideration (e.g. children's reading experiences at a certain grade-level in school). It appears that the types of materials children read—not 'sheer volume'—will largely determine whether and which specialized words will be encountered by children during extensive reading, and whether, or to what extent, those words will be recycled (see Carver 1994, and Carver and Leibert 1995, for related conclusions).

4. While the thematically-related collections share 233 UNQ types with 10+ repetitions in each register, most of which are easily identifiable as being theme-related (*Mummy*, *Mystery*, *Westward Movement*), they only represent 3.5 per cent of the total UNQ shared types. This suggests that the narrative-expository vocabulary distinction cannot be overcome by simply finding narratives and expositories that are related by a common theme.

Furthermore, the control collections in the corpus (4 narratives and 4 expositories *not* related by theme) demonstrate only scant sharing of UNQ types between registers (only *four* UNQ shared types with 10+ repetitions in each register). The important implication here is that the notion of *Wide Reading* may be a valid explanation of how certain higher general frequency

words are acquired (i.e. General High Frequency words that have both high frequency and wide range in basically all reading material), but it appears to be a weak explanation of how large numbers of more specialized (lower general frequency) words are acquired. This is especially true when sheer volume is taken to mean 'any reading' and when such a concept *relies* on the shaky, and apparently unjustified, premise that unqualified reading experiences (i.e. reading for reading's sake) will automatically result in repetitive encounters with numerous unknown words.

In short, it appears that for *Wide Reading* to remain a viable explanation of extensive vocabulary growth through the academic years, proponents must begin to address both register (e.g. narrative vs. expository) and thematic issues (e.g. loosely related vs. tightly related materials) in terms of how these variables affect the *types* of words that children are exposed to, the degree to which those words are actually *recycled*, and thus whether estimates of academic vocabulary growth through reading alone are justified under the current default assumptions of the claim—particularly the notions of incidental acquisition during free pleasure reading of 'good stories'.

5. Even when UNQ shared types appear to be important in both narratives and expositories (i.e. 10+ repetitions in both registers), the actual contexts in which the types are embedded are characteristically different in the two broad registers. In narrative contexts it appears that many of these types are merely incorporated into the larger narrative functions of story telling (plot, characterization, setting, etc.), and they are often semantically impacted by the fictional aspects of the narratives (e.g. a *mummy* chasing people in a Hollywood sense, or a *mummy* in a supernatural sense). Conversely, in the expository contexts it appears that these same types are often highlighted and even explicated, functioning more in the sense of factual terminology (e.g. the processes of making a *mummy*, or the origins of the word *mummy* itself).

In summary, while similar specialized word forms may appear in both narratives and expositories, their acquired meanings by children may be highly dissimilar, depending on the materials in which the words are actually encountered.[17]

## Limitations of the study

It is recognized that certain limitations exist in the design of the current study. These limitations primarily involve the following issues:

1   The linking of children's potential vocabulary growth primarily to issues of word frequency and range in texts rather than actual learner performance, and without consideration of other important variables, both textual (e.g. contextual saliency, syntactic properties of words) and extratextual (e.g. learner motivation, IQ, background knowledge, visual aids).

2   The nonrandom selection of themes and reading materials.

3   The narrow 'type' definition of 'word' to mean unique strings of letters separated by spaces—a definition that does not account for word family relationships, nor for multiple word expressions such as idioms, phrasal verbs, and stock phrases.

It is also recognized that replicative studies using other themes, texts, and controls would be necessary to validate the conclusions made by the researcher. However, it is also felt that the variables examined in this exploratory study are *primary* in terms of certain popular claims made by extensive reading experts concerned about vocabulary, and have generated unique insights regarding children's vocabulary exposure during extensive reading. Furthermore, it is felt that the issues noted above, while limitations in the context of the current study, are also fruitful avenues for future research.


## Looking to the future

The findings in this study suggest that 10- and 11-year-old children may have to employ quite different acquisition processes (i.e. requiring much more specific attention to language) to learn the meanings of a substantial number of words found exclusively, or almost exclusively, in densely-packed expository materials. It is apparent that such words tend not only to be more cognitively challenging and thematically salient than their narrative counterparts, but they are also embedded in contexts that require readers to negotiate a greater proportion of specialized words in order to achieve critical reading comprehension in general—that is, that 95 per cent of words in the surrounding context must already be known before new word learning can take place (Hu and Nation 2000; Laufer 1989, 1997; Liu and Nation 1985; Wixson and Lipson 1991).

The most promising pedagogical approaches for dealing with these issues appear to offer a balance of rich content-area reading experiences coupled with formal instruction aimed at developing students' reading strategies, morphological knowledge, word-analysis strategies, and general study strategies (e.g. Bulgren and Scanlon 1998; Camp 2000; Carlisle 2000; Cunningham 1998; Deshler *et al*. 1996; Guillaume 1998; Guthrie *et al*. 1999; Hulstijn 2001; Loranger 1999; Mahony, Singson, and Mann 2000; Vacca and Vacca 1996; Wesche and Paribakht 2000; Yopp and Yopp 2000). One common component of many of these approaches is an emphasis on graphic organizers as a tool for unpacking or relating difficult content concepts (cf. Mohan 1986, 1990).

The findings in this study also provide confirmatory support to the growing body of studies which have found important linguistic differences between narrative and expository registers (e.g. Biber 1988; Coté, Goldman, and Saul 1998; Hirsh and Nation 1992; Nation and Waring,1997). Perhaps it is time for literacy and vocabulary experts to pay serious attention to a notion of

'narrative and expository macro-genres', recently forwarded by Grabe (2002), and to give more specific consideration to the short-term and long-term ramifications for literacy and vocabulary development that such a notion implies. Could it be possible, for instance, that the oft cited BICS-CALP distinction in ESL contexts (Cummins 1983, 1984, 1989) is not so much a developmental phenomenon as it is a curricular or instructional limitation? In other words, is it possible that an overemphasis on narrative fiction (or, conversely, an underemphasis on expository materials) in the literacy training of ESL learners 'builds in' much of the noted disparities between the development of basic language skills ('easy narratives') and more complex cognitive abilities ('more difficult expositions')?

Furthermore, while effortless 'pleasure' reading has its valuable place, and while it may sound like a desirable goal for literacy experiences and instruction, the conditions for such reading do not always present themselves in real-world contexts, especially in academia and the workplace. In fact, very few individuals make a living by reading stories for pleasure in our current society, but we are often required to negotiate complicated ideas, information, and instructions (Chall 1996; Coulson 1996; Laddo, Reinking, and McKenna 1998; Mikulecky and Kirkley 1998; Snow, Burns, and Griffin 1998; Walberg 1996).

Along these same lines, the vocabulary findings of this study suggest that even fifth grade reading materials, especially those that convey important content information, may require children to possess these same types of sophisticated reading skills. Indeed, the popular notion of uninterrupted 'reading for pleasure', and the popular narrative-centred practices of *Wide Reading*, *Free Reading*, and extensive-reading in general may need to be *balanced* with many more opportunities for students to negotiate the language of more difficult expositions, and to develop conscious strategies for reading the same (cf. Camp 2000; Duke 2000; Fielding and Roller 1992). 'Learning to read by reading' and 'learning from reading' are useful guidelines until the words, and the concepts they represent, get in the way. For many first and second language students, this condition may happen sooner rather than later in their school experiences. By extension, if our attempts to improve or remediate children's reading abilities primarily involve continued exposure to the language in narrative texts, we may actually be contributing to a growing deficit in their academic language abilities which, in turn, may affect their potential for academic success and economic opportunity (cf. Duthie 1994; Moss, Leone, and Dipillo 1997; Pappas 1991).

Perhaps the time has come when we should begin emphasizing the *rewards* of reading with the same vigour as those who have emphasized the pleasure of reading. There is no question that the price for this change in emphasis is more work—not less—on the part of children, as well as those who are willing to assist them. But when, and why, did some of us begin to believe it should be otherwise?

## NOTES

1 One reviewer rightly points out the dangers in relating vocabulary findings without consideration of the different psychological processes employed when L2 learners (of various ages) and L1 learners (of various ages) encounter 'new words' in written contexts (e.g. they may or may not already understand the concepts in their L1; they may or may not already have the words in their spoken vocabularies).

2 It is important to note that Nation's figures in Table 1 are based on word families with each family consisting of a head word (base word) plus many of its inflectional and derivational forms. For example, the family with the head word *absorb* would also include the members *absorbs, absorbed, absorbing, absorption* (see Bauer and Nation 1993, for detailed discussion). Based on this concept, the 3,000 word families are said to represent approximately 5,000 actual word types.

3 Coxhead (2000) has recently developed a new list of academic high frequency words with promising potential. Because the list is new, however, it has not been used as extensively as the UWL in published research.

4 The 95 per cent level is considered a threshold for basic comprehension, which should not be confused with effortless reading. For example, Hirsh and Nation (1992) conclude that approximately 98 per cent of the words (5,000 families or 8,000 actual words) of a given text should be known if the purpose for reading is pleasure, and Nation and Coady (1988) argue that 98 per cent known words is a safer benchmark for learning new vocabulary during reading.

5 Simplified texts have been criticized on the basis that they do not prepare students for the actual readings they will encounter in their academic studies (see Coady 1997a, for review).

6 Ranks were based on the *American Heritage Intermediate* corpus (Carroll, Davies, and Richman 1971).

7 Even the fictional registers themselves exhibit marked differences between each other in their informational characteristics. For instance, science fiction is closest to academic prose in its informational nature, followed by mystery, adventure, and general fiction which tend to group together, and, lastly, romance, which actually exhibits more involved features than informational.

8 The terms *word, words,* and *vocabulary* are used in the study when generic (broad) reference is needed for discussion purposes. The narrower term *type* is used in the study to refer to a written string of letters separated on both sides by spaces. The letters themselves can be upper or lower case or a combination of both. Any change in spelling constitutes a unique type. This form-based definition of *word* corresponds to Level 1 in the Bauer and Nation (1993) taxonomy. *Token* is used in the study to designate raw counts of word types. For instance, if the word *absorb* were to appear 10 times, it would be counted as one *type*, but 10 *tokens*.

9 The researchers indicated that words were chosen from the LOB and Brown corpora if they were high frequency words in at least 10 of the 15 sections (registers) of each corpus (Nation and Hwang 1995).

10 The AHI corpus is 'a computer-assembled selection of 5,088,721 words (*tokens*) drawn in 500-word samples from 1,045 published materials (*texts*). It contains 86,741 different words (*types*)' (Carroll, Davies, and Richman 1971: xiii). The texts were selected 'from 6,162 different titles cited in the response to a national survey of schools in the United States' and 'were chosen and sampled to represent, as nearly as possible, the range of required and recommended reading to which students are exposed in schools grades 3 through 9 in the United States' (p. xiii).

11 It is acknowledged that the cutoff between high frequency and low frequency words in this type of research is somewhat arbitrary. However, according to Nation and Newton (1997), most researchers 'agree that the distinction can be most usefully made somewhere between the most frequent

1,500 words and the most frequent 7,000 words' (p. 239). Therefore, the 2,350 high frequency cutoff used in the current study appears to be justifiable, even though it leans toward the lower end of the range. The figure seems particularly valid given the fact that the texts being analysed in this study are associated with the upper-elementary grades rather than secondary or adult education.

12 The 10+ figure is somewhat liberal based on the research data which indicate a range of 12–20 for vocabulary acquisition during normal reading. However, 10+ is a common benchmark in several published studies and discussions of vocabulary learning through reading (Coady 1997a; Hirsh and Nation 1992; Nation 1990; Wodinsky and Nation 1988).

13 Interestingly, the narrative and expository control collections (not thematically related) share only four 10+ UNQ types between the two four-book collections (*crow, ford, dolphins, creek*). Such a condition does not favour the default vocabulary claims of *Wide Reading*.

14 In the narratives, 76 of the 100 highest frequency types are names of characters.

15 Nation (2001) reported similar trends with adult-level materials.

16 A marked lexical incongruity between the two broad registers occurs even at the more general levels of vocabulary where 43.6 per cent of the AHF types (823 of 1,889) and 20.2 per cent of the GHF types (1,450 of 7,167) are register specific.

17 A reviewer points out that different senses of a word are not necessarily problematical for learning in general. This is based on generative research (Joe 1998), which suggests that new contexts and uses may enhance learning by stretching concepts and providing new associations.

# REFERENCES

Anderson, R. C. 1996. 'Research foundations to support wide reading' in V. Greaney (ed.): *Promoting Reading in Developing Countries*. Newark, DE: International Reading Association, pp. 55–77.

Anderson, R. C. and W. E. Nagy. 1992. 'The vocabulary conundrum,' *American Educator* Winter: 14–19, 44–47.

Bauer, L. and P. Nation. 1993. 'Word families,' *International Journal of Lexicography* 6/4: 253–79.

Biber, D. 1988. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.

Brinton, D., M. A. Snow, and M. Wesche. 1989. *Content-based Second Language Instruction*. New York: Newbury House.

Bulgren, J., and D. Scanlon. 1998. 'Instructional routines and learning strategies that promote understanding of content area concepts,' *Journal of Adolescent & Adult Literacy* 41/4: 292–302.

Camp, D. 2000. 'It takes two: Teaching with twin texts of fact and fiction,' *The Reading Teacher* 53/5: 400–8.

Campion, M. E. and W. B. Elley. 1971. *An Academic Vocabulary List*. Wellington: NZCER.

Carlisle, J. F. 2000. 'Awareness of the structure and meaning of morphologically complex words: Impact on reading,' *Reading and Writing: An Interdisciplinary Journal* 12: 169–90.

Carrell, P. L. and J. G. Carson. 1997. 'Extensive and intensive reading in an EAP setting,' *English for Specific Purposes* 16/1: 47–60.

Carroll, J. B., P. Davies, and B. Richman. 1971. *The American Heritage Word Frequency Book*. New York: Houghton Mifflin.

Carver, R. P. 1994. 'Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: Implications for instruction,' *Journal of Reading Behavior* 26/4: 413–37.

Carver, R. P. and R. W. Leibert. 1995. 'The effect of reading library books at different levels of difficulty upon gain in reading ability,' *Reading Research Quarterly* 30/1: 26–8.

Chall, J. S. 1996. 'American reading achievement: Should we worry?' *Research in the Teaching of English* 30/3: 303–10.

Chamot, A. U. and M. O'Malley. 1987. 'The cognitive academic language learning approach: A bridge to the mainstream,' *TESOL Quarterly* 21/2: 227–47.

Cho, K. and **S. D. Krashen.** 1994. 'Acquisition of vocabulary from the Sweet Valley Kids series: Adult ESL acquisition,' *Journal of Reading* 37/8: 662–7.

Coady, J. 1993. 'Research on ESL/EFL vocabulary acquisition: Putting it in context' in T. Huckin, M. Haynes, and J. Coady (eds): *Second Language Reading and Vocabulary Learning*. Norwood, NJ: Ablex, pp. 3–23.

Coady, J. 1997a. 'L2 vocabulary acquisition through extensive reading' in J. Coady and T. Huckin (eds): *Second Language Vocabulary Acquisition*. Cambridge: Cambridge University Press, pp. 225–37.

Coady, J. 1997b. 'L2 vocabulary acquisition: A synthesis of the research' in J. Coady and T. Huckin (eds): *Second Language Vocabulary Acquisition*. Cambridge: Cambridge University Press, pp. 273–90.

Coady, J., J. Magoto, P. Hubbard, J. Graney, and K. Mokhtari. 1993. 'High frequency vocabulary and reading proficiency in ESL readers' in T. Huckin, M. Haynes, and J. Coady (eds): *Second Language Reading and Vocabulary Learning*. Norwood, NJ: Ablex, pp. 217–28.

Coté, N., S. R. Goldman, and E. U. Saul. 1998. 'Students making sense of informational text: Relations between processing and representation,' *Discourse Processes* 25/1: 1–53.

Coulson, A. J. 1996. 'Schooling and literacy over time: The rising cost of stagnation and decline,' *Research in the Teaching of English* 30/3: 311–27.

Coxhead, A. 2000. 'A new academic word list,' *TESOL Quarterly* 34/2: 213–38.

Crandall, J. 1993. 'Content-centered learning in the United States,' *Annual Review of Applied Linguistics* 13: 111–26.

Cummins, J. 1983. 'Conceptual and linguistic foundations of language assessment' in S. S. Seidner (ed.): *Issues of Language Assessment: Vol. 2. Language Assessment and Curriculum Planning*. Wheaton, MD: National Clearinghouse for Bilingual Education, pp. 7–16.

Cummins, J. 1984. *Bilingualism and Special Education: Issues in Assessment and Pedagogy*. Clevedon, Avon: Multilingual Matters.

Cummins, J. 1989. *Empowering Minority Students*. Sacramento: California Association for Bilingual Education.

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

Deshler, D. D., E. S. Ellis, and B. K. Lenz. (eds) 1996. *Teaching Adolescents with Learning Disabilities: Strategies and methods* 2nd edn. Denver: Love Publishing Company.

Duke, N. K. 2000. '3.6 minutes per day: The scarcity of informational texts in first grade,' *Reading Research Quarterly* 35/2: 202–24.

Duthie, C. 1994. 'Nonfiction: A genre study for the primary classroom,' *Language Arts* 71: 588–95.

Elley, W. B. 1991. 'Acquiring literacy in a second language: The effect of book-based programs,' *Language Learning* 41/3: 375–411.

Elley, W. B. 1996. 'Using book floods to raise literacy levels in developing countries' in V. Greaney (ed.): *Promoting Reading in Developing Countries*. Newark, Delaware: International Reading Association, pp. 148–62.

Elley, W. B. and F. Mangubhai. 1983. 'Impact of reading on second language learning,' *Reading Research Quarterly* 19/1: 53–67.

Ellis, R. and X. He 1999. 'The roles of modified input and output in the incidental acquisition of word meanings,' *Studies in Second Language Acquisition* 21: 175–80.

Ellis, R., Y. Tanaka, and A. Yamazaki. 1994. 'Classroom interaction, comprehension, and the acquisition of L2 word meanings,' *Language Learning* 44/3: 449–91.

Enright, D. S. and M. McCloskey. 1988. *Integrating English*. Reading, MA: Addison-Wesley.

Fielding, L. and C. Roller. 1992. 'Making difficult books accessible and easy books acceptable,' *The Reading Teacher* 45/9: 678–85.

Francis, W. N. and H. Kucera. 1982. *Frequency Analysis of English Usage*. Boston: Houghton Mifflin.

Freyd, P. and J. Baron. 1982. 'Individual differences in the acquisition of derivational morphology,' *Journal of Verbal Learning and Verbal Behavior* 21: 282–95.

Gardner, D. 1997. Vocabulary Acquisition in Theme-based Instruction: Assessing the lexical load of theme-based reading materials in the language arts. Unpublished dissertation pilot study, Northern Arizona University.

Gardner, D. 1999. Vocabulary acquisition through reading: Assessing the lexical composition of theme-based text collections in upper-elementary education. Doctoral dis-

sertation, Northern Arizona University *Digital Dissertations*, AAT9937445.

Gibbons, P. 1998. 'Classroom talk and the learning of new registers in a second language,' *Language and Education* 12/2: 99–118.

Guillaume, A. M. 1998. 'Learning with text in the primary grades,' *The Reading Teacher* 51/6: 476–86.

Guthrie, J. T., E. Anderson, S. Alao, and J. Rinehart. 1999. 'Influences of concept-oriented reading instruction on strategy use and conceptual learning from text,' *The Elementary School Journal* 99/4: 343–66.

Grabe, W. 1991. 'Current developments in second language reading research,' *TESOL Quarterly* 25/3: 375–405.

Grabe, W. 2002. 'Narrative and expository macro-genres' in A. Johns (ed.): *Genre in the Classroom: Multiple perspectives*. Mahwah, NJ: LEA. pp. 249–67.

Grabe, W. and F. Stoller. 1997. 'Content-based instruction: Research foundations' in M. A. Snow and D. M. Brinton (eds), *The Content-based Classroom: Perspectives on Integrating Language and Content*. White Plains, NY: Addison Wesley Longman, pp. 5–21.

Hayes, D. P. and Ahrens, M. 1988. 'Vocabulary simplification for children: A special case of ''motherese''?' *Journal of Child Language* 15: 395–410.

Hazenberg, S. and J. H. Hulstijn. 1996. 'Defining a minimal receptive second-language vocabulary for non-native university students: An empirical investigation,' *Applied Linguistics* 17/2: 145–63.

Hirsh, D. and P. Nation. 1992. 'What vocabulary size is needed to read unsimplified texts for pleasure?' *Reading in a Foreign Language* 8/2: 689–96.

Hu, M. and I. S. P. Nation. 2000. 'Vocabulary density and reading comprehension,' *Reading in a Foreign Language* 13/1: 403–30.

Hulstijn, J. 2001. 'Intentional and incidental second language learning: A reappraisal of elaboration, rehearsal and automaticity' in P. Robinson (ed.): *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press. pp. 258–86.

Hwang, K. and P. Nation. 1989. 'Reducing the vocabulary load and encouraging vocabulary learning through reading newspapers,' *Reading in a Foreign Language* 6/1: 323–35.

Joe, A. 1998. 'What effects do text-based tasks promoting generation have on incidental vocabulary acquisition?' *Applied Linguistics* 19/3: 357–77.

Johansson, S. 1978. *Manual of Information to Accompany the Lancaster-Oslo/Bergen corpus of British English, for use with digital computers*. Department of English, University of Oslo.

Kinsella, K. 1997. 'Moving from comprehensible input to ''learning to learn'' in content-based instruction' in M. A. Snow and D. M. Brinton (eds): *The Content-based Classroom: Perspectives on Integrating Language and Content*. White Plains, NY: Addison Wesley Longman, pp. 46–68.

Krashen, S. 1982. *Principles and Practices in Second Language Acquisition*. New York: Pergamon Press.

Krashen, S. 1985. *The Input Hypothesis: Issues and implications*. New York: Longman.

Krashen, S. 1989. 'We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis,' *The Modern Language Journal* 73/4: 440–64.

Krashen, S. 1993a. 'The case for free voluntary reading,' *The Canadian Modern Language Review* 50/1: 72–82.

Krashen, S. 1993b. *The Power of Reading: Insights from the research*. Englewood, CO: Libraries Unlimited.

Laddo, L. D., D. Reinking, and M. C. McKenna. 1998. 'Technology and literacy education in the next century: Exploring the connection between work and schooling,' *Peabody Journal of Education* 73/3&4: 273–89.

Landauer, T. K. and S. T. Dumais. 1997. 'A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge,' *Psychological Review* 104/2: 211–40.

Laufer, B. 1989. 'What percentage of text-lexis is essential for comprehension?' in C. Lauren and M. Nordman (eds): *Special Language: From Humans Thinking to Thinking Machines*. Philadelphia: Multilingual Matters, pp. 316–23.

Laufer, B. 1997. 'The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess' in J. Coady and T. Huckin (eds): *Second Language Vocabulary Acquisition*. Cambridge: Cambridge University Press, pp. 20–34.

Laufer, B. and P. Nation. 1995. 'Vocabulary

size and use: Lexical richness in L2 written production,' *Applied Linguistics* 16/3: 307–22.

**Leong, C. K.** 2000. 'Rapid processing of base and derived forms of words and grades 4, 5 and 6 children's spelling,' *Reading and Writing: An Interdisciplinary Journal* 12: 277–302.

**Liu, N.** and **I. S. P. Nation.** 1985. 'Factors affecting guessing vocabulary in context,' *RELC Journal* 16/1: 33–42.

**Loranger, A. L.** 1999. 'The challenge of content area literacy: A middle school case study,' *The Clearing House* 72/4: 239–43.

**Mahony, D., M. Singson,** and **V. Mann.** 2000. 'Reading ability and sensitivity to morphological relations,' *Reading and Writing: An Interdisciplinary Journal* 12: 191–218.

**Meara, P.** 1997. 'Towards a new approach to modeling vocabulary acquisition' in N. Schmitt and M. McCarthy (eds): *Vocabulary: Description, Acquisition and Pedagogy.* Cambridge: Cambridge University Press, pp. 109–21.

**Mikulecky, L.,** and **J. R. Kirkley.** 1998. 'Literacy instruction for the 21st-century workplace,' *Peabody Journal of Education* 73/3&4: 290–316.

**Miller, G.** and **P. Gildea.** 1987. 'How children learn words,' *Scientific American* 257/3: 94–9.

**Mohan, B.** 1986. *Language and Content.* Reading, MA: Addison-Wesley.

**Mohan, B.** 1990. 'LEP students and the integration of language and content: Knowledge structures and tasks' in C. Simich-Dudgeon (ed.): *Proceedings of the First Research Symposium on Limited English Proficient Students' Issues.* Washington, DC: Office of Bilingual Education & Minority Language Affairs, pp. 113–60.

**Moss, B., S. Leone,** and **M. Dipillo.** 1997. 'Exploring the literature of fact: Linking reading and writing through information trade books,' *Language Arts* 74: 418–29.

**Nagy, W. E.** 1988. *Teaching Vocabulary to Improve Reading Comprehension.* Newark, Delaware: International Reading Association.

**Nagy, W.** 1997. 'On the role of context in first- and second-language vocabulary learning' in N. Schmitt and M. McCarthy (eds): *Vocabulary: Description, Acquisition and Pedagogy.* Cambridge: Cambridge University Press, pp. 64–83.

**Nagy, W. E.** and **R. C. Anderson.** 1984. 'How many words are there in printed school English?' *Reading Research Quarterly* 19/3: 304–30.

**Nagy, W. E., R.C. Anderson,** and **P. Herman.** 1987. 'Learning word meanings from context during normal reading', *American Educational Research Journal* 24: 237–70.

**Nagy, W. E., I. N. Diakidoy,** and **R. C. Anderson.** 1993. 'The acquisition of morphology: Learning the contributions of suffixes to the meanings of derivatives,' *Journal of Reading Behavior* 25/2: 155–80.

**Nagy, W. E.** and **P. A. Herman.** 1984. *Limitations of Vocabulary Instruction. (Tech. Rep. No. 326).* Champaign, IL: University of Illinois Center for the Study of Reading.

**Nagy, W. E.** and **P. A. Herman.** 1985. 'Incidental vs. instructional approaches to increasing reading vocabulary,' *Educational Perspectives* 23: 16–21.

**Nagy, W. E.** and **P. A. Herman.** 1987. 'Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction' in M. G. McKeown and M. E. Curtis (eds): *The Nature of Vocabulary Acquisition.* Hillsdale, NJ: Lawrence Erlbaum, pp. 19–35.

**Nagy, W. E., P. A. Herman,** and **R. C. Anderson.** 1985. 'Learning words from context,' *Reading Research Quarterly* 20/2: 233–53.

**Nation, I. S. P.** 1990. *Teaching and Learning Vocabulary.* New York: Newbury House.

**Nation, I. S. P.** 1993. 'Vocabulary size, growth, and use' in R. Schreuder and B. Weltens (eds): *The Bilingual Lexicon.* Amsterdam: John Benjamins, pp. 115–34.

**Nation, P.** 1997. 'The language learning benefits of extensive reading,' *The Language Teacher* 21/5: 13–16.

**Nation, I. S. P.** 2001. *Learning Vocabulary in Another Language.* Cambridge: Cambridge University Press.

**Nation, P.** and **J. Coady.** 1988. 'Vocabulary and reading' in R. Carter and M. McCarthy (eds): *Vocabulary and Language Teaching.* New York: Longman, pp. 97–110.

**Nation, P.,** and **K. Hwang.** 1995. 'Where would general service vocabulary stop and special purposes vocabulary begin?' *System* 23/1: 35–41.

**Nation, P.** and **J. Newton.** 1997. 'Teaching vocabulary' in J. Coady and T. Huckin (eds): *Second Language Vocabulary Acquisition.* Cambridge: Cambridge University Press, pp. 238–54.

Nation, P. and R. Waring. 1997. 'Vocabulary size, text coverage and word lists' in N. Schmitt and M. McCarthy (eds): *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press, pp. 6–19.

Pappas, C. 1991. 'Fostering full access to literacy by including information books,' *Language Arts* 68: 449–62.

Paribakht, T. S. and M. Wesche. 1999. 'Reading and ''incidental'' L2 vocabulary acquisition: An introspective study of lexical inferencing,' *Studies in Second Language Acquisition* 21: 195–224.

Raptis, H. 1997. 'Is second language reading vocabulary best learned by reading?' *The Canadian Modern Language Review* 53/3: 566–80.

Scarcella, R. and C. Zimmerman. 1998. 'Academic words and gender: ESL student performance on a test of academic lexicon,' *Studies in Second Language Acquisition* 20: 27–49.

Schmitt, N. and R. Carter. 2000. 'The lexial advantages of narrow reading for second language learners,' *TESOL Journal* Spring 2000: 4–9.

Snow, C. E., M. S. Burns, and P. Griffin. (eds) 1998. *Preventing Reading Difficulties in Young Children*. Washington, DC: National Academy Press.

Stanovich, K. E., R. F. West, A. E. Cunningham, J. Cipielewski, and S. Siddiqui. 1996. 'The role of inadequate print exposure as a determinant of reading comprehension problems' in C. Cornoldi and J. Oakhill (eds): *Reading Comprehension Difficulties: Processes and Intervention*. Mahwah, New Jersey: Lawrence Erlbaum, pp. 15–32.

Stoller, F. L. and W. Grabe. 1997. 'A six-T's approach to content-based instruction' in M. A. Snow and D. M. Brinton (eds): *The Content-based Classroom: Perspectives on Integrating Language and Content*. White Plains, NY: Addison Wesley Longman, pp. 78–94.

Sutarsyah, C., P. Nation, and G. Kennedy. 1994. 'How useful is EAP vocabulary for ESP? A corpus based case study,' *RELC Journal* 25/2: 34–50.

Tyler, A. and W. Nagy. 1989. 'The acquisition of derivational morphology,' *Journal of Memory and Language* 28: 649–67.

Vacca, R. T. and J. L. Vacca. 1996. *Content Area Reading* 5th edn. New York: Harper Collins.

Walberg, H. J. 1996. 'U.S. schools teach reading least productively,' *Research in the Teaching of English* 30/3: 328–43.

Walmsley, S. A. 1994. *Children Exploring their World: Theme Teaching in the Elementary Schools*. Portsmouth, NH: Heinemann.

Wesche, M. B. and T. S. Paribakht. 2000. 'Reading-based exercises in second language vocabulary learning: An introspective study,' *The Modern Language Journal* 84/2: 196–213.

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

Wixson, K. K. and M. Y. Lipson. 1991. *Reading Disability*. New York: Harper Collins.

Wodinsky, M. and P. Nation. 1988. 'Learning from graded readers,' *Reading in a Foreign Language* 5/1: 155–61.

Xue, G. and P. Nation. 1984. 'A university word list,' *Language Learning and Communication* 3/2: 215–29.

Yopp, R. H. and H. K. Yopp. 2000. 'Sharing informational text with young children,' *The Reading Teacher* 53/5: 410–23.

Zimmerman, C. B. 1997. 'Historical trends in second language vocabulary instruction' in J. Coady and T. Huckin (eds): *Second Language Vocabulary Acquisition*. New York and Cambridge: Cambridge University Press, pp. 5–19.

## APPENDIX A: REFERENCES OF CHILDREN'S BOOKS IN CORPUS

### Books in thematic collections

### Mystery narrative multiauthor

Brenner, B. 1972. *Mystery of the plumed serpent* (rev. edn). New York: Alfred A. Knopf.

Elmore, P. 1992. *Susannah and the Purple Mongoose Mystery*. New York: Dutton Children's Books.

Konigsburg, E. L. 1967. *From the Mixed-up Files of Mrs. Basil E. Frankweiler*. New York: Atheneum.

Wortis, A. 1991. *Windcatcher*. New York: Avon Books.

## Mystery narrative uniauthor

Kehret, P. 1995a. *Frightmares: Cat Burglar on the Prowl*. New York: Pocket Books.

Kehret, P. 1995b. *Frightmares: Don't Go Near Mrs. Tallie*. New York: Pocket Books.

Kehret, P. 1996a. *Frightmares: Backstage Fright*. New York: Pocket Books.

Kehret, P. 1996b. *Frightmares: Screaming Eagles*. New York: Pocket Books.

## Mystery expository multiauthor

Beattie, O. and J. Geiger. 1992. *Buried in Ice*. New York: Scholastic.

Bisel, S. C. 1990. *The Secrets of Vesuvius*. New York: Scholastic.

Jackson, D. M. 1996. *The Bone Detectives*. Boston: Little, Brown and Company.

Sheely, R. 1993. *Police Lab: Using Science to Solve Crimes*. Silver Moon Press.

## Mystery expository uniauthor

Simon, S. 1976. *Ghosts*. Philadelphia: J.B. Lippincott.

Simon, S. 1979. *Creature from Lost Worlds*. New York: J.B. Lippincott.

Simon, S. 1981. *Mad Scientists, Weird Doctors, and Time Travelers in Movies, TV, and Books*. New York: J.B. Lippincott.

Simon, S. 1997. *Strange Mysteries from Around the World* (rev. edn). New York: Morrow Junior Books.

## Mummy narrative multiauthor

Masterman-Smith, V. 1982. *The Great Egyptian Heist*. New York: Four Winds Press.

Peck, R. 1986. *Blossom Culp and the Sleep of Death*. New York: Bantam Doubleday Dell.

Snyder, Z. K. 1967. *The Egypt Game*. New York: Bantam Doubleday Dell.

Voight, C. 1991. *The Vandemark Mummy*. New York: Fawcett Juniper.

## Mummy narrative uniauthor

Bellairs, J. 1978. *The Treasure of Alpheus Winterborn*. New York: Harcourt Brace Jovanovich.

Bellairs, J. 1983. *The Mummy, the Will, and the Crypt*. New York: Dial Books for Young Readers.

Bellairs, J. 1984. *The Dark Secret of Weatherend*. New York: Dial Books for Young Readers.

Bellairs, J. 1996. *The Curse of the Blue Figurine* (rev. edn). New York: Puffin Books.

## Mummy expository multiauthor

Bendick, J. 1989. *Egyptian Tombs*. New York: Franklin Watts.

Lauber, P. 1985. *Tales Mummies Tell*. New York: Thomas Y. Crowell.

Putnam, J. 1993. *Eyewitness Books: Mummy*. New York: Alfred A. Knopf.

Wilcox, C. 1993. *Mummies and their Mysteries*. Minneapolis: Carolrhoda Books.

## Mummy expository uniauthor

Millard, A. 1982. *Ancient Egypt*. London: Granada.

Millard, A. 1987. *Great Civilizations: Egypt 3118 BC–AD 642*. New York: Franklin Watts.

Millard, A. 1995. *Mysteries of the Pyramids*. Brookfield, CT: Copper Beech Books.

Millard, A. 1996. *Pyramids*. New York: Kingfisher.

## Westward movement narrative multiauthor

Conrad, P. 1985. *Prairie Songs*. New York: Harper & Row, Publishers.

Fleischman, S. 1988. *By the Great Horn Spoon* (rev. edn). Boston: Little, Brown and Company.

Lawlor, L. 1986. *Addie Across the Prairie*. Niles, IL: Albert Whitman & Company.

Moeri, L. 1994. *Save Queen of Sheba* (rev. edn). New York: Puffin Books.

## Westward movement narrative uniauthor

Wilder, L. I. 1971a. *Little House in the Big Woods* (rev. edn). New York: Harper & Row.

Wilder, L. I. 1971b. *Little House on the Prairie* (rev. edn). New York: Harper & Row.

Wilder, L. I. 1971c. *Farmer Boy* (rev. edn). New York: Harper & Row.

Wilder, L. I. 1971d. *On the Banks of Plum Creek* (rev. edn). New York: Harper & Row.

## Westward movement expository multiauthor

Blumberg, R. 1989. *The Great American Gold r.u.s.h.* New York: Bradbury Press.

Freedman, R. 1983. *Children of the Wild West.* New York: Scholastic.

Sandler, M. W. 1994. *Cowboys.* New York: Harper Collins.

Tunis, E. 1961. *Frontier Living* (chapters 9–19). Cleveland: The World Publishing Company.

## Westward movement expository uniauthor

Hakim, J. 1994a. *A History of US (book five): Liberty for All* (chapters 1–9). New York: Oxford University Press.

Hakim, J. 1994b. *A History of US (book five): Liberty for All* (chapters 10–18). New York: Oxford University Press.

Hakim, J. 1994c. *A History of US (book five): Liberty for All* (chapters 19–27). New York: Oxford University Press.

Hakim, J. 1994d. *A History of US (book five): Liberty for All* (chapters 28–36). New York: Oxford University Press.

## Books in nonthematic collections

### Narrative control collection

L'Engle, M. 1962. *A Wrinkle in Time.* New York: Farrar, Straus, and Giroux.

O'Dell, S. 1960. *Island of the Blue Dolphins.* New York: Dell Publishing.

Paterson, K. 1977. *Bridge to Terabithia.* New York: Harper & Row.

Raskin, E. 1978. *The Westing Game.* New York: Viking Penguin.

### Expository control collection

Dow, L. 1990. *Whales: A Great Creatures of the World Book.* New York: Weldon Owen Pty Limited.

Heinrichs, A. 1992. *America the Beautiful: Montana* (2nd ed.) Chicago: Children's Press.

Maestro, B. and G. Maestro. 1996. *The Voice of the People: American Democracy in Action.* New York: Lothrop, Less & Shepard Books.

Ride, S. and S. Okie. 1986. *To Space and Back.* Orlando, FL: Harcourt Brace.