

Vocabulary Analysis of a Graded Reader

Andy W. Fuller

Colorado State University

### Abstract

An important component of SLA research is L2 vocabulary development and its interrelation with other aspects of L2 competency. Many studies have focused on vocabulary knowledge relative to levels of comprehension of authentic materials and academic texts. As reported by Nation (2001), vocabulary studies have shown that participants need to know 95%- 98% of the running words in a text to “gain adequate comprehension” (p. 147). A recent study involving 661 participants conducted by, Grabe, Jiang, and Schmitt (in press) showed that there was a relatively linear correlation between reading comprehension and vocabulary knowledge in a given text. In other words, the more vocabulary you know, the better you can comprehend a text. Although the above studies involved academic or authentic texts, one could assume that vocabulary coverage and comprehension remain constant for lower level learners as well. As beginners often use graded readers, one was chosen for this project. This paper will examine the vocabulary of a beginner graded reader to gain insight into adapted texts for ESL/EFL students.

*Keywords:* vocabulary, graded readers, frequency, coverage

### Vocabulary Analysis of a Graded Reader

Reading is an important aspect in learning a second language; therefore, many researchers have focused on vocabulary knowledge and its effects on a reader's comprehension of a text. Vocabulary studies have shown that the more words a learner knows in a text, the higher the level of comprehension that is possible for the reader (Nation, 2001; Grabe, Jiang, & Schmitt, in press). While these studies reported somewhat different results, there is a strong indication that knowing less than 95% of the vocabulary in a text can severely hinder reading comprehension. Nation (2001) explained that 80 % vocabulary coverage in a text means that for a language learner, "one word in every five is unknown", 90 % coverage means "one word in every ten is unknown", 95% coverage means "one word in every twenty is unknown", and 98-99% means there is "one unknown word in every 50-100 running words" (p. 147); therefore, below the 95% coverage level, little comprehension occurs on the part of the reader.

According to Nation (2001), language learners who know the 2000 high frequency word families on the *General Service List* (GSL) will know 87.4% of the tokens in most works of fiction (Table 1.7, p.17). This number is much lower than the minimum 95% of coverage needed by language learners to read and comprehend a text. As this percent is based on a large corpus of un-adapted fiction texts, how can students who are still well below the first 2000 word families read within the 95%-98% coverage as part of their English program? One method of controlling the vocabulary input for students is by using graded readers in the classroom.

Graded readers are usually works of fiction or novels that "have been prepared so that they stay within a strictly limited vocabulary" (Nation, 2001, p.162). For example, graded readers can be written to only contain the first 1000 words on the GSL. Learners can read a few graded readers at this level and then move into readers that contain words from the second 1000

words on the GSL. According to Nation (2001), learners moving from one level to the next will often encounter many new words. He suggested a focus on key vocabulary to help students as they transition from one level to the next. One method of focusing on key vocabulary is through glossing. A gloss is a short definition or explanation of a potentially difficult word or phrase contained in the graded reader. According to Nation (2001), glosses are most helpful in the margins because students can comprehend the text without severely interrupting the reading process. A suggested sequence of a graded reader scheme is offered by Nation in the table below.

Level	Number of word families
1	500
2	700
3	1,000
4	1,500
5	2,000
6	3,000
7	5,000

(2001, p. 166)

Graded readers that contain glosses are widely used in ESL programs. Therefore, it is important for ESL teachers to know which words are glossed and how many words the students will need to know to have 95% or more coverage of the text. The purpose of this paper is to analyze the text and glosses contained in *Knights of the Round Table* and examine word frequency and coverage relative to the intended beginner audience.

### **Corpus selection**

The graded reader *Knights of the Round Table* was selected for this project. This book is part of the *High Point* series for ESL learners created by Hampton-Brown publishing and is listed as a level two beginning graded reader for 2<sup>nd</sup>-4<sup>th</sup> grade students. According to Hampton-Brown (2002) they “used the most current scientifically based research in the development of the

intervention program *High Point* designed for struggling readers and English...” (p.2). The *High Point* series is also based on TESOL standards of language proficiency. Because this series is based on research, it seemed an appropriate candidate for a frequency and vocabulary profile analysis. Roughly half of the book (4,200 words) was scanned into a text format and then analyzed using lextutor.

### **Coverage and Frequencies**

As *Knights of the Round Table* is a graded reader, it seemed appropriate to separate the text and the glosses for the coverage and frequency analysis. This was done to see if the glosses were written in easier language than the text they are meant to explain. In terms of text coverage, *Knights of the Round Table* contained a total of 539 word families with 84% of those word families being in the first 1000 in the GSL. Only 5.38 % of the word families belong to the second 1000 words families in the GSL. Low frequency word families make up 10% of the text and only three of the word types are from the academic word list (*stable, finally, invisible*).

The glosses contained a total of 172 word families with 88.5% of those families from the first 1000 in the GSL. A total of 6.5% of the word families belonged to the second 1000 words in the GSL, a slightly higher number than in the actual text. However, low frequency word families only made up 5% of the words in the glosses. These figures are clearly illustrated in the following table.

Table 1.1 *Knights of the Round Table* (1985) vocabulary profile

Levels	Knights of the Round Table (text)	Knights of the Round Table (glosses)
1 <sup>st</sup> 1000	84.28% (410*)	88.50% (150 *)
2 <sup>nd</sup> 1000	5.38% (129*)	6.50% (22 *)
Academic	0.12% (+)	0%
Other	10.22%	5%

\*word families

+Academic- stable, finally, invisible

Low frequency items in the text and glosses were divided into three separate categories: related low frequency types, other low frequency types, and proper nouns. Related types were defined as belonging to the genre of medieval fiction. Other types were seen as general low frequency words that could have been used in different genres. Proper nouns included any names or places used in the story. Further examination of all low frequency items revealed that a large number of the actual tokens of low frequency words were in the proper noun category. For example, Arthur showed up 85 times in the text, whereas armor and wizard showed up only 4 times. A sample listing of these categories is offered below.

Table 1.2 Low frequency word families

Related low frequency types (text)	Other low frequency types (text)	Proper nouns (text)
armor, dragon, knight, mist, monster, prancing, silvery, swirl, witch, wizard, tale, throne, vanished, magic, maidens	afterward, blasted, blushed, hissing, halfway, muttering, jutting, sighed, gentlest.	Arthur, Merlin, Ector, England, Morgan, Guinevere, Kay, Camelot

*Knights of the Round Table* (1985)

The low frequency words in the glosses were divided into the same categories as the low frequency words in the text. Examination of low frequency items in the glosses showed that many of them were proper nouns like Kay, Arthur, Ector, and Merlin. The related group of low frequency words contained some of the same words as the related group in the actual text.

However, when these words were used in the glosses, they had been previously explained in the text. For example, in *Knights of the Round Table*, the word *mist* is explained as *light rain and clouds*. Shortly after this gloss, *In its place* is explained as *where the mist had been, now*.

According to Nation (2001), language learners need to meet new vocabulary words multiple times to actually learn them. As stated above, *Knights of the Round Table* used previously explained words to clarify new words or phrases. This could create an environment where the reader meets unfamiliar words more than once. The more times a reader meets unfamiliar vocabulary in context, the more likely that person is to actually learn the new vocabulary.

### **Extensive or Intensive Reading**

Also important in examining the percent of words known relative to a text is the type of reading done in an L2: intensive or extensive reading. Intensive reading involves a specific focus on key vocabulary, focused input, and may increase a learner's breadth of vocabulary knowledge, whereas extensive reading is largely geared towards fluency development, incidental learning, and may enhance a learner's depth of vocabulary knowledge (Nation, 2001).

What this means in terms of using *Knights of the Round Table* in the classroom is quite simple; students who are well below the 1000 word family level would benefit from an intensive reading scheme designed around the book. Pre-reading exercises that build background knowledge for the medieval genre and focused input on important vocabulary would increase the ability of the student to comprehend the text and learn new high frequency items. According to Nation, "native speaking children beginning school already know close to 5,000 word families..." (2001, p. 156). However, an ESL student in second grade will most likely not have

such a strong command of the English language and will need focused instruction to learn high frequency words from the text.

Students who are just under the 2000 word family level could use *Knights of the Round Table* and other graded readers at the same level for extensive reading purposes. An ESL teacher with many different levels and genres of graded readers would be well equipped to include extensive reading in their curriculum.

### **Conclusion**

This paper examined the word frequency and vocabulary coverage of a beginner graded reader *Knights of the Round Table*. The intended beginner audience seems appropriate as the book only contains 539 word families with over 400 of these word families being from the first 1000 families on the GSL. Also, the glosses provide a significant jump in coverage from 89.66% to 95% of the GSL. No true collocations were prevalent in the text. I assume that because the text was adapted for low level learners, it does not always reflect authentic language. There were a number of low frequency words in the text. However, many of these low frequency words were proper nouns and medieval genre specific words that could be easily explained by pointing to pictures in the book. In sum, this book would be valuable in an intensive or extensive reading scheme for beginning learners of the appropriate demographic.



## References

- Grabe, Jiang, Scmitt. (in press). The Percentage of Words Known in a Text and Reading Comprehension. *Modern Language Journal*, 94. Retrieved from <http://www.cal.nau.edu/english/faculty/grabe.asp>
- Hampton-Brown. (2002). *The Research Basis of High Point*. Retrieved from [http://www.ngsp.com/Portals/0/downloads/Research\\_Base\\_HP.pdf](http://www.ngsp.com/Portals/0/downloads/Research_Base_HP.pdf)
- Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. United Kingdom: Cambridge University Press.

Appendix 1

<b>Levels</b>	<b>Fiction</b>
1 <sup>st</sup> 1000	82.3%
2 <sup>nd</sup> 1000	5.1%
Academic	1.7%
Other	10.9%

Nation (2001, p. 17)

<b>Level</b>	<b>Number of word families</b>
1	500
2	700
3	1,000
4	1,500
5	2,000
6	3,000
7	5,000

Nation (2001, p. 166)

<b>Levels</b>	<b>Knights of the Round Table (text)</b>	<b>Knights of the Round Table (glosses)</b>
1 <sup>st</sup> 1000	84.28% (410*)	88.50% (150 *)
2 <sup>nd</sup> 1000	5.38% (129*)	6.50% (22 *)
Academic	0.12% (+)	0%
Other	10.22%	5%

\*word families

+Academic- stable, finally, invisible

<b>Related low frequency types (text)</b>	<b>Other low frequency types (text)</b>	<b>Proper nouns (text)</b>
armor, dragon, knight, mist, monster, prancing, silvery, swirl, witch, wizard, tale, throne, vanished, magic, maidens	afterward, blasted, blushed, hissing, halfway, muttering, jutting, sighed, gentlest	Arthur, Merlin, Ector, England, Morgan, Guinevere, Kay, Camelot, Uther

*Knights of the Round Table* (1985)

<b>Related low frequency types (glosses)</b>	<b>Other low frequency types (glosses)</b>	<b>Proper nouns (Glosses)</b>
Dragon, knight, magic, mist, witch	Churchyard, contest, hooray	Kay, Arthur, Ector, Merlin

*Knights of the Round Table* (1985)

<b>Type of reading</b>	<b>Learning goals</b>	<b>% vocabulary coverage</b>
Intensive reading	Developing Language Developing strategy use knowledge	Less than 95% coverage
Extensive reading for language growth	Incidental vocabulary learning Reading skills	95-98% coverage
Extensive reading for fluency development	Reading quickly	99-100% coverage

Nation (2001, p. 150)