

Shared Identities: Our Interweaving Threads



Customizing reading texts using software

Mark Alberding
Rikkyo University

Reference data:

Alberding, M. (2009). Customizing reading texts using software. In A. M. Stoke (Ed.), *JALT2008 Conference Proceedings*. Tokyo: JALT.

This paper explains how teachers can use readily-available software such as a text editor, a word-processing program, and the free vocabulary-analysis program *RANGE*, to adapt a text and make it a more suitable match for their students' vocabulary levels. First, the difficulty of a text as measured by its lexical content is discussed and a brief overview of adapting texts is provided. Next, an example text provides material for a discussion of adaptation choices. Detailed instructions for the software procedures described in the paper are likewise included.

この論文では、教師が容易に利用できるテキストエディター、ワードプロセッシングプログラム、無料語彙分析プログラムRANGEのようなソフトウェアを利用し、いかに学習者の語彙レベルに順応したものを作成するかを説明する。始めに、語彙により難易度を計られたテキストの難しさ、そして作成したテキストの基本的側面について簡潔に述べ、次に、テキスト例をあげ、教材を適応させるためのいくつかの手段を提示する。この論文の中で出てくるソフトウェアの手順に関する細かな指示も含まれていることをここに記載する。

Readings class teachers may find themselves wanting or being required to use authentic materials, texts written for an audience of first-language readers that have “not been specifically produced for the purpose of language teaching” (Nunan, 1989, p. 54). However, for all but the most advanced learners, such texts are too difficult to read comfortably. One solution is to adapt authentic texts to make them more accessible to second-language learners, that is, to make them easier to read and comfortably understand. Although there are different approaches to simplification, some of which are briefly discussed below, this paper focuses on a single approach: adjusting the difficulty of vocabulary items in the text by using simpler synonyms and marginal glosses. This approach, which is suitable for adapting texts for any level of learner, is carried out using readily available software that can be used by anyone who has a basic working knowledge of computers. Only three programs are required, with the first two being standard on almost any modern computer: a text editor, such as Notepad (Windows) or TextEdit (Macintosh), and a word processing program such as Microsoft Word. The third program-RANGE-is a free program designed

by Paul Nation. It is powerful yet simple to use and can be downloaded from his website. Appendix 3 of this paper contains the necessary web address and gives step-by-step instructions for the major portions of the process outlined later in the paper.

Vocabulary and text difficulty

Unknown vocabulary is undoubtedly a significant barrier to reading comprehension in a second language. In order to read with “adequate comprehension,” the threshold at which comprehension is interfered with by unknown vocabulary (Nation, 2006, p. 61), learners need to have a minimum vocabulary level. Waring and Nation (1997) estimated a knowledge of 3000-5000 words to be necessary for initial access to authentic texts. Laufer (1997) suggests a minimum level of 3000 words. Additionally, one needs to know how much coverage of different kinds of text is provided by a vocabulary of a given level. Hirsh and Nation (1992) found a 5000 word level vocabulary was necessary to achieve approximately 97% coverage of a selection of novels. Examining newspaper content, Nation (2006) found that to achieve 98% coverage a vocabulary of 8000-9000 words was required, while 95% coverage occurred with around 4000 words. However, it appears that few if any second-language learners can gain adequate comprehension of non-fiction writing even with 98% coverage, according to Kurnia (as cited in Nation, 2006).

The numbers presented above are a selective glance at text difficulty based on lexical content but they illustrate the challenge of using authentic articles: the texts are likely beyond all but the most advanced learners. The author’s

experience, including some testing over the years, leads him to believe that many of his students have a vocabulary of 2000-3000 words on average. According to Nation’s (2006) analysis, a 2000 word vocabulary provides less than 83% coverage of newspaper articles. This means that almost 1 in 20 words will be unintelligible. Neither adequate comprehension nor fluent reading, the two main goals the author has for his classes, will occur in these circumstances.

Adapting texts: an overview

One approach to making authentic texts more accessible to learners with a limited vocabulary is to adapt the texts. The variety of methods for adapting texts and the theories behind the methods are beyond the scope of this paper but it is useful to distinguish between simplification, in which difficult words and grammatical constructs are replaced with simpler ones, and elaboration, which involves such things as parenthetical paraphrases, noun phrase appositives, and glossing difficult words. In general, research has pointed to elaboration as superior to simplification (Oh, 2001; Yano, Long, & Ross, 1994). However, Young (1999) points out that the complexity and sheer quantity of variables involved in the reading process may be the reason for inconsistent results in studies on the effectiveness of text simplification and Leow (1997) notes that part of the problem of such studies is varying definitions of what constitutes simplification and the lack of a systematic approach. In the author’s approach, simplification means substituting a word with an easier-to-understand synonym or near synonym, or rephrasing a phrase or clause with words that are easier to understand, such as words that occur more frequently according to word lists.

One aspect of elaboration is vocabulary glosses, which have frequently been found effective in making texts more accessible to second language learners. Roby's (1999) literature-review article notes evidence of the effectiveness of reader access to glosses during reading. Ko (2005) found comprehension advantages for second-language readers using glosses. In particular, marginal glosses are seen to encourage a more fluid reading style while aiding understanding. Hulstijn, Hollander, and Greidanus (1996) recommended marginal glosses as effective because they reduced the need for learners to use their dictionaries while reading, which "interferes with the process of constructing a mental representation of text meaning" (p. 337). Marginal glosses were found by Davis (1989) to facilitate comprehension and lead to more fluent reading by second language learners and they were found to be more effective for vocabulary learning than explanatory appositives by Watanabe (1997). Nation (2006) notes their value in adapting texts to bridge the gap between graded readers and unsimplified text. Additionally, when asked, learners have indicated a preference for marginal glosses as opposed to other types of glosses such as on separate pages, in the text itself, or at the top or bottom of the page (Ko, 2005; Jacobs, Dufon, & Hong, 1994).

The degree and type of text adaptation will be determined by the learners' vocabulary knowledge, the goals the instructor has for the adapted material, and practical considerations such as the amount of time that can be devoted to adapting texts. In the author's case, the goal was to enable students to read authentic texts more fluently and at a level of comprehension that would allow the reading to

be pleasurable rather than burdensome, and would lead to a greater likelihood of success with comprehension activities and ensuing discussions. Given the amount of time available for adapting texts in addition to other teaching duties, the author has found that a combination of simplification of select words and phrases or clauses, and elaboration through marginal glosses has served his pedagogical purposes.

Texts can be adapted for any level of learner, with the limiting factor being the amount of vocabulary learners at a given level possess. If learners have only a 500 word vocabulary a much greater degree of simplification will be required than if they have a 4000 word vocabulary. Therefore, it is necessary to have some idea of learners' vocabulary knowledge so that informed decisions can be made when simplifying texts. This can come from the results of a vocabulary levels test or a less objective but still useful indicator such as learners' self-reported level of comfortable reading using graded readers, which state the vocabulary level for a given title. The author currently adapts texts mainly for first-year students with an estimated vocabulary of around 2000 words. This estimate was derived from the average results of a vocabulary levels test which had been administered to the same first-year student population in previous years (ideally each class of each year would be tested but this is no longer practicable in the author's case.)

The process of adapting a text discussed in this paper is intended to make authentic texts accessible to learners with limited vocabularies, when teachers use such texts in class through choice or because of curricular requirements. Of course, a text can no longer be concerned "authentic" once its language has been altered for language teaching purposes.

Once adapted, the text becomes a language learning text. In fact, this is the author's goal, as the purpose of the procedure is to make available to teachers and learners content that is not available "off the shelf" at the learners' proficiency level. This includes current events articles from newspapers or magazines, op-ed essays, writing on topics related to the learners' humanities majors and excerpts of fiction or non-fiction books, to name just a few. In short, almost any kind of written text that a teacher would want to use can be adapted through simplification.

Adapting a text to make it more accessible to learners

This section of the paper provides an overview of the process of adapting a text, a brief look at the *RANGE* program used to analyze texts, and a sample text which illustrates the author's approach to adapting authentic texts. In the author's experience, the time required to adapt a text varies from 30 minutes to 1 hour or longer depending on the length and complexity of the original, and whether the text requires editing for length or content such as removing irrelevant passages or shortening paragraphs. Obviously, a longer, more lexically-complex text will take greater time to adapt than a shorter, more lexically-simple text as more decisions about simplification will be necessary.

Overview of the process

1. Locate the text you wish to adapt. The internet is an ideal source since many newspapers and magazines have their content available online, although one needs

to be cognizant of copyright and use restrictions. One option is using copyright-free articles from websites set up to provide such content without charge; the example used in this paper comes from such a site. If one wishes to use a print source, this must be scanned into a digital file using a scanner with optical character recognition software (a step which is not covered in this paper).

2. Edit the text for length, irrelevant or unwanted passages, etc.
3. Create a document using the edited text.
4. Analyze the vocabulary of the edited text using the *RANGE* program.
5. Adapt the text, by simplifying selected words and phrases and creating a gloss in the document left margin for other words and phrases.

The *RANGE* program

The vocabulary analysis program *RANGE*, designed by vocabulary expert Paul Nation, is central to the adaptation process described here. The program is easy to use and comes with word lists against which the lexical content of a text is compared. Two sets of lists are available: the *GSL* and *AWL* set and the *BNC* set. The former consists of the first two thousand words of English from *A General Service List of English Words* (GSL) by Michael West and the *Academic Word List* (AWL) by Averil Cox, a list of 570 words derived from a corpus of academic texts and which do not occur in the first 2000 words of the GSL; the latter consists of word lists created using the spoken section of the British National

Corpus (BNC). The example in this paper was created using the former set of lists.

Sample text

The text used as an example here-*Survey Highlights the Importance of Teaching Children Good Social Skills*-comes from the copyright-free article website ARAContent (<http://www.aracontent.com>).

Lexical content and coverage

Table 1 shows information about the vocabulary content of the article obtained from *RANGE*. Focusing on types, unique occurrences of words (i.e. not including repetitions of the same word), we see that approximately 79% of the text is covered by the first two thousand words of the GSL. The remainder of text coverage is split almost evenly between the AWL and words not in the lists. It is this 21% that is likely to interfere with achieving adequate comprehension which, as seen earlier, isn't likely to occur much below the 98% level.

Text marking

A feature of *RANGE* useful for adapting texts is the *mark words* option, which marks the words in a text according to the word list in which the words appear. Figure 1 shows a portion of text after processing with *RANGE* using word marking. The full marked text appears in Appendix 2.

Table 1. Tokens, types and text coverage by different word lists

Word list	Tokens / text coverage (%)	Types* / text coverage (%)
GSL 1st 1000 words	184 / 74.2	96 / 71.1
GSL 2nd 1000 words	25 / 10.1	10 / 7.4
Academic word list	18 / 7.3	14 / 10.4
Not in any lists	21 / 8.5	15 / 11.1
Total	248	135

*excluding proper nouns and names

4-year-old<!> children may already know how to tie<2> their own shoelaces<!> and spell<2> their first and last names. But as preschool<!> looms<!> around the corner<2>, parents<2> are worried<2> how well their children will fit in with the rest of the classroom<!>.

According to a nationwide survey<3> conducted<3> of 1,000 parents<2> by Mom Central Inc<!>., the majority<3> of parents<2> feel the same way with 90 percent<3> considering social skills<2> to be vital<!> to their children's happiness and confidence<2>.

Key:
 no marking: GSL first 1000 words of English
 <2>: GSL second 1000 words of English
 <3> :AWL
 <!> :not in any of the lists.

Figure 1: Text marked by RANGE

Making adjustments based on the RANGE results

The marked text enables instructors to quickly identify vocabulary that may be outside of the learner's knowledge and can be simplified or glossed to make it more accessible.

As mentioned earlier, the learners for whom the author prepares texts were considered to have a vocabulary knowledge which includes the first 2000 words of English. Accordingly words in this range were generally ignored. However, nearly 21% of types remained which needed either simplifying or glossing to make the text more accessible. A summary of the changes made appears in Table 2. Some of these changes are discussed below but a complete discussion of all changes isn't possible here; readers may compare the marked text in Appendix 2 with the final version distributed to students in Appendix 1 to see all changes made by the author.

Table 2. Summary of adaptations made to original text

Word list	Types*	Types unchanged	Types simplified	Types glossed
GSL 1st 1000 words	96	96	1	1
GSL 2nd 1000 words	10	8	1	1
Academic word list	14	6	0	8
Not in any lists	15	5	7	3
Total	135	115	9	13

*excluding proper nouns and names

Twenty types occurring in the AWL or not in any list were simplified or glossed, which was assumed to make them

more accessible to learners. In principle, the approach taken was that words that were not in any lists were simplified and the words from the AWL were glossed. The rationale is that words not in any lists are likely to be of such low frequency that they are not worth expending energy on learning explicitly, as through glosses, while the AWL words may be of higher frequency and encountered more often and hence a gloss may be a learning aid. Words that were simplified were cross-checked against the word lists to make sure they occurred in the first 2000 words. Before adaptation, approximately 21% of the text was covered by words from the AWL and those not in any lists. After adaptation, this number fell to 6%, an improvement of 15%. However, the total text coverage accessible to the learners may be even greater yet than the apparent post-simplification/gloss figure of 94% because the 11 types comprising the remaining 6% were left unchanged precisely because they were judged by the author to already be known to the learners. These included words which fell outside the lists such as classroom and kids, and some which occurred in the AWL, such as area and team. Of course, the process of judging which words are known to learners is a subjective one. Any two teachers may reach a different result in this regard and it is possible that some words judged to be known to learners are in fact not known; thus the qualification that the coverage *may* be greater than 94% in this case.

In fact, the process of simplifying a text requires decisions informed by experience at many turns as the software is identifying words only as they appear in standardized lists and cannot account for all aspects of a word's difficulty, such as those attributable to collocation or idiomatic usage. For

example, in this case one word from the second 1000 words of the GSL-*critical*-was glossed because of its perceived difficulty in the context of the article, and two from the first 1000 words which occurred in collocations-*more* and *raised*-were, respectively, glossed or simplified. Also, some words not found in any lists were glossed as it was judged that they could not be simplified while maintaining authenticity or that the results of simplification might appear stilted. Finally, some words and phrases required a more involved approach. For example *but as preschool looms around the corner* contains two words not in any list-*preschool* and *looms*-and one from the second 1000 words-*corner*. It was not possible to simplify the former without changing the latter so the entire clause was changed to *will soon enter preschool*. Similarly, *executive* is marked as not in the lists, but in the text is a collocate in the title *chief executive officer* which clearly needs to be glossed as unit. As powerful as the software is in analyzing a text, teachers will need to interpret the results in order to make the best adaptations to texts.

Conclusion

The purpose of this paper was to show how teachers can use readily-available tools to adapt authentic texts for use by learners with limited vocabularies. The author accepts the assumption that simplifying or glossing difficult words improves reader accessibility to texts. Using the vocabulary analysis program *RANGE*, teachers can make informed decisions about which items to simplify or gloss rather than relying solely on intuition, which can be misleading. The amount of simplification depends on the vocabulary level of the target audience and the goals of the instructor but

the procedure can be used for texts destined for learners at different proficiency levels. It is hoped that teachers will find the procedure of value when circumstances require the use of authentic texts beyond the reach of their learners.

Mark Alberding holds an MA in English (TESOL) from San Francisco State University. He is currently a *kyoiku kyoshi* lecturer at Rikkyo University in Tokyo. His interests extend to whatever makes his teaching more effective for his students.

References

- Davis, J. N. (1989). Facilitating effects of marginal glosses on foreign language reading. *Modern Language Journal*, 73(1), 41-48.
- Hirsh, D., & Nation, I. S. P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8(2), 689-696.
- Hulstijn, J., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *Modern Language Journal*, 80(3), 327-339.
- Jacobs, G. M., Dufon, P., & Fong, C. H. (1994). L1 and L2 vocabulary glosses in L2 reading passages: Their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading*, 17(1), 19-28.

- Ko, M. H. (2005). Glosses, comprehension, and strategy use. *Reading in a Foreign Language*, 17. Retrieved November 12, 2008, from <http://nflrc.hawaii.edu/rfl/October2005/ko/ko.html>.
- Laufer, B. (1997). What's in a word that makes it hard or easy: Some intralexical factors that affect the learning of words. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 140-155). Cambridge: Cambridge University Press.
- Leow, R. P. (1997). Simplification and second language acquisition. *World Englishes*, 16(2), 291-296.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59-82.
- Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge: Cambridge University Press.
- Oh, S-Y. (2001). Two types of input modification and EFL reading comprehension: Simplification versus elaboration. *TESOL Quarterly*, 35(1), 69-96.
- Roby, W. B. (1999). What's in a gloss? *Language Learning and Technology*, 2(2). Retrieved November 12, 2008, from <http://llt.msu.edu/vol2num2/roby/>.
- Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition*, 19(3), 287-307.
- Waring, R., & Nation, I. S. P. (1997). Vocabulary size, text coverage, and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6-19). Cambridge: Cambridge University Press.
- Yano, Y., Long, M. H., & Ross, S. (1994). The effects of simplified and elaborated texts on foreign language comprehension. *Language Learning*, 44(2), 189-219.
- Young, D. J. (1999). Linguistic simplification of SL reading material: Effective instructional practice? *Modern Language Journal*, 83(3), 350-366.

Appendix 1

Excerpt of a formatted, adapted text given to students

Click here to view Appendix 1

Appendix 2

Marked text

4-year-old<1> children may already know how to tie<2> their own shoelaces<1> and spell<2> their first and last names. But as preschool<1> looms<1> around the corner<2>, parents<2> are worried<2> how well their children will fit in with the rest of the classroom<1>.

According to a nationwide survey<3> conducted<3> of 1,000 parents<2> by Mom Central Inc<1>., the majority<3> of parents<2> feel the same way with 90 percent<3> considering social skills<2> to be vital<1> to their children's happiness and confidence<2>.

Nearly eight out of 10 parents<2> also think social skills<2> are more important than academic<3> skills<2> when it comes to their child's overall<3> happiness. As a matter of

fact, parents gave social skills a higher ranking than academic skills on the survey in nearly every area of child development.

“More than ever, our children must get along with others to function effectively,” says Stacy DeBroff, chief executive officer of Mom Central. “In this age of team sports and structured play, it has never been more critical for our children to master socialization skills. From children’s play groups to collaboration in the classroom, kids today engage in significantly more structured group activities, raising the profile and the necessity for good social skills.”

According to the survey, one in five parents feel overwhelmed with teaching social skills and more than one-third say that teaching social skills leads to frustration. In response, Stacy DeBroff has developed some tips parents can use to help their child learn social skills in a positive and reinforcing way.

Appendix 3

Instructions for adapting authentic texts

Once an article is selected, the time for the adaptation procedure varies from 30 minutes to 1 hour or more depending on the length and complexity of the original, and whether the text requires editing for length or content such as removing irrelevant passages.

What you need to adapt texts

The following software is needed to adapt texts as described in this article:

- A text editor program that creates plain text files in ASCII format (e.g. Windows’ *Notepad*, Apple’s *TextEdit*).
- A word processing program that creates document files that can be formatted (e.g. *Microsoft Word*, *Word Perfect*, *Open Office Writer*.)
- The free *RANGE* vocabulary-analysis program. The program can be downloaded from Nation’s webpage on the University of Victoria website at <http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx>. The program only runs under Windows operating systems, unfortunately.

If one is expecting to adapt texts on a regular basis, it is desirable to create a document template so that one can quickly and easily create adapted texts without having to re-create the final document format each time. The format the

author has found most useful appears below in outline form and can be seen in a document form in Appendix 1.

The format for an A4-sized page is as follows:

Top margin 45 mm

Bottom margin: 22 mm

Left margin 81 mm

Right margin 22 mm

Header: 6 mm from edge

Footer: n/a

Header information		
		Text box for article title and source information
Text box for vocabulary gloss		Article (paste text then add line numbers)

Once the desired format has been created, save the result as a template (choose the *template* option when saving the document in Microsoft Word). This makes the format always available simply by choosing *File>New>Templates-On my computer...* when creating a new document. Please note that the *step-by-step instructions* below assume that a template has been created.

Step-by-step instructions

Step 1: Copy and paste the text of your selected article into a plain text file. If there are portions of the article you wish to remove because of overall length, lack of relevancy, etc., make those edits at this point.

Step 2: Create a new Word document using your article document template. Save/name the document (which will be referred to as *article document* from this point forward).

Step 3: Copy the text from the plain text file and paste it into the article document.

Step 4: Close the text file, noting its location so you can easily access it later.

Step 5: Open the *RANGE* program and load the plain text file you created in step 1.

(Instructions for *RANGE* accompany the program and are not duplicated here).

Step 6: Run *RANGE*, making sure to check the *Mark Texts* box.

Step 7: Use the resulting *.mrk* file to analyze the vocabulary content of your article. Note that double-clicking on the *.mrk* file will not open it: Windows will display a dialog box that states *Windows cannot open this file*. However, the *.mrk* file is simply a plain text file with a different extension and you can open it using the plain text editor you used for other steps above. Alternately, you can change the extension of the file from *.mrk* to *.txt* so that in the future you can open the file simply by double clicking it.

Step 8: Make the desired adjustments to the text of your article document, which can include simplifying and/or glossing difficult vocabulary.