

## A defence of simplification

I S P NATION – Victoria University of Wellington, New Zealand  
JEAN PAUL DEWEERDT – Free University, Belgium

### ABSTRACT

This paper argues that simplified or graded readers are an essential part of a language learning program if learners of all proficiency levels are to have the opportunity to do incidental language learning through reading, and to develop fluency in reading. Unsimplified texts do not allow for this kind of learning at beginning and intermediate levels because they contain too great a density of unknown words and too many different unknown words. Evidence is provided to support this from a corpus study of versions of *Dinotia*. Many criticisms of simplified texts apply only to poorly simplified texts and to the poor use of such texts in curriculum planning.

### Introduction

Graded readers are books specially written for learners of English using a controlled vocabulary and grammar. A typical graded reader series consists of books written at five or six vocabulary levels, beginning at around 300–400 words and increasing in stages to around 2500 words. The Oxford Bookworms series, for example, has six levels, as shown in Table 1.

Table 1: The Oxford Bookworms series

Level	New words at each level	Cumulative total words	Range of length in running words	Examples of titles
1	400	400	4743–5890	<i>White Death, Mutiny on the Bounty</i>
2	300	700	5511–7960	<i>Dinotia, Dead Man's Island</i>
3	300	1000	8819–12 194	<i>The Bronte Story, Love Story</i>
4	400	1400	14 342–20 142	<i>Lord Jim, Reflex</i>
5	400	1800	20 379–25 272	<i>Heat and Dust, I Robot</i>
6	700	2500	24 840–31 501	<i>Decline and Fall, Meteor</i>

A graded reader may be simplified from a text for native speakers, such as *Jane Eyre*, *Dinotia* or *Lord Jim*, or may be a book written solely for second language learners, such as *White Death*, *One Way Ticket* or *The Mystery of Allegria*.

Some teachers and researchers feel that there is something wrong with texts simplified from literary works in particular, but also more generally with any

text that involves controlled and simplified language use. This uneasiness about graded readers may be expressed either indirectly, through not using them in extensive reading programs as in the Elley and Mangubhai (1981) study, or directly through criticism of the books and their use. Yano, Long and Ross (1994), for example, criticise simplified texts for being unauthentic, inhibiting comprehension and providing poor conditions for learning. The purpose of this article is to show that graded readers are an essential part of a second or foreign language learning program. This will be done firstly by examining the criticisms of graded readers, and secondly by showing the value of graded readers using a comparison of a simplified version and the original unsimplified version of *Dynavla*.

### Criticisms of simplified texts

Graded readers that are simplified from literary texts written for native speakers are criticised for being unauthentic (Honeyfield 1977). A similar criticism is applied to all texts that are written within a deliberately restricted vocabulary. There is no doubt that reading a simplified version of *Dynavla* is not the same as reading the original, but no writer of a graded reader would suggest that it is. There is still the feeling, however, that a text which has been altered in the drastic way that a simplified text has is no longer an authentic text. This criticism depends on the meaning of 'authentic'. 'The meaning of 'authentic' that is being used in the criticism is 'written for native speakers'. In a very perceptive article, Widdowson (1976) pointed out that 'authentic' or 'authenticity' can be viewed in a different way. Authenticity is not a characteristic of texts, but is the result of the interaction between a reader and a text. If a learner reads a text, and responds to it in a way that we might expect of someone who comprehends the text, then reading the text is authentic for that learner. This response might involve understanding the text, enjoying its message, seeing the strengths and weaknesses in its content and expression, or seeing its contribution to a wider field. As we shall see later, a text that is authentic, in that it was produced for native speakers, may be too difficult for learners to respond to in an authentic way. This view of authenticity is similar to the modern view of validity. A test is not valid in its own right, but is valid when it is used for the purpose for which it was designed and examined.

Simplified texts are also criticised because the restriction on writers to use short, simple sentences can result in choppy and unnatural discourse (Honeyfield 1977; Yano et al 1994), and may result in poor cohesive reference and an overreliance on implicit, rather than explicit, conjunction relationships. These can make the texts difficult to comprehend. These criticisms may be true of poorly written simplifications, but there are many excellent simplifications

that are a joy to read. Hill (in Day and Bamford 1998) provides a very useful and substantial list of these high quality texts. It is unfair and misleading to condemn simplifications as a whole because some are poorly done.

Graded readers are also criticised because they provide poor conditions for learning (Yano et al 1994). These criticisms include the following:

- 1 Reading graded readers that are pitched lower than the learners' level can lower the quality of their output.
- 2 Removing difficult vocabulary denies learners access to what they need to learn.
- 3 Reading texts with little unknown vocabulary discourages the development of generalisable coping skills, such as guessing from context and dictionary use.

These are all criticisms of the ways in which graded readers are fitted into a language course. In the remainder of this article, we will look at how the use of graded readers needs to be managed through the matching of the level of graded readers to learners' proficiency levels, and to the learning goals of the various strands of a course. It should then be clear that the criticisms listed above are really criticisms of poor syllabus design, rather than criticisms of graded readers themselves. Let us now look at the various parts of a language course to see how graded readers can fit into these parts, where unsimplified texts would not be suitable.

### The strands of a language course

One way of ensuring that there is a balance of appropriate learning activities in a course is to see a well-balanced course as containing a roughly equal balance of the four strands of meaning-focused input, language-focused learning, meaning-focused output and fluency development.

The strand of meaning-focused input involves incidental learning through listening and speaking where around 98 per cent of the running words are already familiar or pose no learning burden to the learners (Hu and Nation 2000). In this strand, learners' focus is on the message of the texts. Graded readers have a very important part to play in this strand of a course, both as sources of listening input as well as reading input, particularly for learners at the beginning and intermediate levels of proficiency. Much of the justification for the meaning-focused input strand of a course comes partly from the work of Krashen (1985) and advocates of the comprehension approach (Nord 1980), and from research on 'book floods' (Elley 1991). The book flood experiments involve dramatically increasing the amount of reading that language learners

do, and observing their increase in proficiency over a range of language skills and aspects of language knowledge. Elley (1989) has also shown that listening to stories can be a source of vocabulary learning. The conditions for learning from input involve being able to comprehend the input, even though it contains a few features that are just beyond the learners' knowledge. The input should be interesting and should engage the learners in following the message; it contains. Advocates of learning from input (Krashen 1985; Elley 1991) want to see it occupy the vast majority of the learning time. For reasons explained below, it seems wiser to see it occupying about one-quarter of the learning time.

The strand of language-focused learning involves the deliberate study of pronunciation practice. This strand includes vocabulary study, grammar activities, pronunciation practice and intensive reading where learners work with difficult, often unsimplified, texts. There is now considerable research (Ellis 1994; Spada 1997; Long and Robinson 1998) which shows that deliberate intentional learning has positive benefits in language learning and learning from input, partly through consciousness raising. Research in vocabulary learning (Nation 2001: 263–316) has shown that intentional learning produces faster and better results than incidental learning. It is important in course design, however, that meaning-focused learning and language-focused learning are not seen as competitors, but rather, are seen as complementing each other. The language-focused learning strand (Ellis 1994 calls it formal instruction and form-focused learning) should not occupy too much class time at the expense of more message-focused learning, and so it seems useful to limit it to around 25 per cent of the learning time.

The strand of meaning-focused output involves incidental learning through speaking and writing. Learners' focus is on communicating messages. The justification for a meaning-focused output strand has a theoretical basis in the work of Swain (1985) and experimental support from research such as that of Joe (1998) and Newton (1995), which show that learning can occur, and is enriched, when learners produce language. Swain argues that having to produce messages makes the learners look at language in a new way. Language features, particularly grammatical elements, that were given little attention in listening and reading, gain a greater importance in production when learners are pushed to make choices in how to say something.

The strand of fluency development does not involve new language items but focuses on helping learners become fluent with what they already know. This fluency development is meaning-focused and needs to occur across the four skills of listening, speaking, reading and writing. Simplified texts have an important role to play in this strand because they are the best means of providing material that contains no unknown items. There has been increasing interest in fluency development (Schmidt 1992; Skehan 1996), particularly with findings

that fluency increases are typically accompanied by improvements in accuracy and complexity of language use (Arehart and Nation 1991). As well as improving access to language features, fluency practice – in the form of speed reading and repeated reading, timed speaking and writing activities, and extensive listening activities – can enrich and consolidate knowledge of language features. Although fluency increases seem to transfer well between languages (West 1941; Cramer 1975), there seems to be a need to provide fluency practice in each of the four skills of listening, speaking, reading and writing. Fast speakers are not necessarily fast readers. Fluency improvement best occurs when learners work with material that contains no unknown language features, where there is some pressure to perform at a higher than usual rate, where there is quantity of practice, and where the focus is on comprehending or producing messages. Fluency development tends to be neglected in most language courses, but learners need it in order to make the best use of what they already know. It is thus advisable to explicitly devote about 25 per cent of learning time to fluency development.

Giving roughly equal time to the four strands is an arbitrary decision aimed largely at counteracting spending too much time on direct language teaching and counteracting the neglect of fluency development and learning from truly comprehensible input and output.

In the following comparison of a simplified and an unsimplified text, we will focus on extensive reading in the two strands of meaning-focused input and fluency development, because it is in these two areas particularly that simplified material in the form of graded readers is essential.

#### Learning through meaning-focused input

Learning through meaning-focused input involves comprehension and conditions such as repetition that encourage learning. Table 2 lists some of the conditions affecting comprehension and learning. Three texts are analysed in the table – the Oxford Bookworms version of *Dinastia* (a level 2 simplified text of 7957 running words), the first 7965 running words of the unsimplified original of *Dinastia*, which is of roughly the same length as the simplified version to control for the effects of text length, and the complete unsimplified original of *Dinastia* (161 952 running words).

Whenever the phrase 'the first 2000 words' or a similar phrase is used in this article, it refers to the 2000 word families of Michael West's (1953) *A general service list of English words*. Many of these 2000 words do not occur at all in the simplified and unsimplified versions of *Dinastia* used in this study. West's list is now old and in some ways out of date. Unfortunately, in spite of enormous advances in computing and the development of corpora, no comparable carefully made list is available that is suitable for children or young

adults. Research on West's list (Nation and Hwang 1995) indicates that the list still works well and editing the list brings about only very small changes in coverage. In spite of this, a replacement is long overdue.

**Table 2: Conditions for comprehension and learning in a simplified version and the original version of *Dracula***

Conditions	Simplification	Short original	Complete original
Percentage of words in first 2000 or proper nouns	98.6%	91.3%	92.8%
Total number of different word families	556	1416	5640
Word families not in first 2000 or not proper nouns	30 (5.4%)	530 (37.4%)	3822 (67.8%)
Repetitions of word types not in first 2000 or not proper nouns	1 × 7 2 × 2 3 × 4	1 × 391 2 × 46 3 × 16	1 × 3039 2 × 785 3 × 320

Word families (Bauer and Nation 1993) have been used for most of the counting in this study. A word family includes the stem form of a word, its inflected forms and its regularly related, largely transparent, derived forms. For example, the word family *mend* includes the following members – *mends*, *menda*, *menting*, *mendable*, *mender*, *unmendable*. Once learners are familiar with the inflectional system of English and the most frequent, regular and productive derivational affixes, the word family is the most suitable unit for counting the receptive vocabulary load of texts. Otherwise, *mend*, *mends*, *menting* etc would be counted as different words.

Research on vocabulary density and text comprehension (Hu and Nation 2000) indicates that for fiction texts, ideally, 98 per cent of the running words in the text should be familiar or should pose no comprehension burden for the readers, if the text is to be read without consulting a dictionary. The following list of different coverage densities and average number of ten-word lines per unknown word shows the importance of coverage.

- 98% coverage one unknown word in every five lines
- 95% coverage one unknown word in every two lines
- 90% coverage one unknown word in every line

Table 2 shows that, for learners with a vocabulary of 2000 words or less, the simplified text (98.6 per cent coverage) would have 1 word outside the first 2000 words in every 7 lines. This provides a very supportive context for each unknown word. The short unsimplified text (91.3 per cent coverage) would have 1 word outside the first 2000 in every 1.2 lines. This is a very

small context for the unknown words. Appendix 1 has an example of the simplified and unsimplified texts with words outside the first 2000 marked in bold to show what this is like in practice.

The density of unknown words is one measure of difficulty. A related measure is the actual number of different unknown words outside the first 2000 and that are not proper nouns. This affects difficulty because each new word is another item to guess or look up, and if there are many of these, then comprehension is difficult to achieve. Related to this are the repetitions of these unknown words. If a reasonable proportion of the coverage of the running words outside the first 2000 is made up of repeated words, then these have a very good chance of being learned, and will quickly become known words that then add to the coverage of known words for the remainder of the text. The words *vampire*, *wolves*, *howl* and *coffin* are examples of repeated words outside the first 2000 in *Dracula*.

Table 2 shows that there are only 30 word families outside the first 2000 and not proper nouns in the simplified text, and several of these words are repeated many times. In the short unsimplified text, on the other hand, 530 of the total 1416 word families (37.4 per cent) are not in the first 2000. In the complete unsimplified version, this rises to 67.8 per cent of the word families. Moreover, in the short unsimplified text, 391 of the 530 word families occur only once. This means that each one is potentially a comprehension and learning problem. Thus, there are around 530 possible interruptions in the short unsimplified text and relatively few in the simplified text. These 'one-timers' make up a significant proportion of the total different words in the unsimplified text, and this proportion increases with the length of the text. These one-timers require effort for comprehension, and this effort is not repaid by the opportunity to meet them again when they would be less of a burden and would have a chance of being learned.

Even in the simplified text there are many words, particularly within the first 2000 words, that occur only once in the text. However, because these words are from the most frequent 2000 words of English, they are worth learning. It is a reasonable piece of advice to give to a learner of English that any word in a graded reader is worth learning. This piece of advice is certainly not true for an unsimplified text. Here are some of the uncommon words in the short unsimplified text from *Dracula* – *alacrity*, *aquiline*, *bayting*, *crags*, *diligence* (a type of stagecoach), *engendered*, *gaitre*, *hospitars*, *oleander*, *polyglot*. Some of these words will eventually be useful additions to a broad and rich vocabulary, but there are thousands of more useful words to learn before these.

We have seen that unsimplified text can place a very heavy vocabulary burden on second language readers. Unsimplified text has too many of its running words outside most second language learners' vocabulary knowledge, so that

readers with a limited vocabulary meet an unknown word in every few running words. These unknown words make up a very large group of words, most of which occur only once in the text. A large number of texts would need to be read before many of them were met again. For learners of English with a vocabulary smaller than 2000 words, most unsimplified text is just too difficult and does not provide the conditions necessary for learning through meaning-focused input. Simplified texts, as in graded readers, can provide these conditions and are thus essential if there is to be a substantial meaning-focused input strand to a course. Struggling with difficult text can be a useful part of a course, but it is part of the language-focused learning strand.

### Fluency development

Fluency development requires practice across the four skills of listening, speaking, reading and writing. The material used to develop fluency should contain no unknown vocabulary, so that learners can focus on making the best and most fluent use of what they already know. For learners with a vocabulary of less than 2000 words, unsimplified text is completely unsuitable because it will not only contain unknown words, but will contain large numbers of them. To develop fluency in reading, learners should be working with graded readers that are well within their proficiency level. They need to read many of them, and they need to push themselves to read as fast as they can. Nation and Wang (1999) have shown that reading several graded readers provides excellent opportunities to meet the same vocabulary many times. Reading unsimplified text does not provide suitable conditions for fluency development for either beginning or intermediate learners. Research on developing speed reading with language learners (West 1941; Carter 1975) has shown that, with the use of simplified material, substantial increases in speed are possible within reasonably short periods of time. As most second language learners read at speeds far below a normal untrained reading rate of around 250 words per minute, developing fluency in reading should be a major goal of most English courses. It is a goal that can be easily achieved with the use of simplified material.

### Other novels

This article has compared the controlled text of a graded reader with uncontrolled unsimplified text. It might be argued that a novel such as *Dracula* is an unusual kind of text in that it was written over 100 years ago and contains a lot of foreign vocabulary because of its setting in a non-English-speaking country. Table 3 presents the coverage figures, the number of word types not found in the most frequent 2000 words of English, and the percentage of one-timers in a range of novels, to show that *Dracula* compares with other unsimplified

fiction texts and that the arguments presented here about the unsuitability of unsimplified texts for learning from meaning-focused input and fluency development are generally true.

Table 3: Vocabulary statistics for a range of unsimplified fiction texts

	<i>Dracula</i>	<i>39 Steps</i>	<i>Washington Square</i>	<i>Something New</i>
Length	161 952	41 640	64 672	75 973
Percentage coverage by the first 2000 words of English	92.8%	89.3%	89.7%	86.1%
Total word types	9640	5347	5929	8079
Number of word types outside first 2000 occurring only once, including proper nouns	3039	1753	1951	2799

Table 3 contains data largely from novels which usually provide the most favourable vocabulary load for learners, while newspapers and academic texts provide a much heavier vocabulary load.

### Conclusion

Unsimplified texts do have important roles to play in language courses, and, for most learners, being able to comprehend them represents a major goal of a language course. But the end is not usefully the means. Good pedagogy involves helping learners reach their goals through suitably staged steps. In the beginning and intermediate stages of language learning, controlled texts are among the most suitable means of bringing the important strands of learning from meaning-focused input and fluency development into play. When learners' vocabulary is large enough, then unsimplified texts can fill these roles. At the beginning and intermediate stages, unsimplified texts may have a role to play in intensive reading and strategy development in the language-focused learning strand of the course.

Learners of English are extremely fortunate in that there is this tremendous resource of hundreds of graded readers at a variety of levels. This resource should be used as fully as possible and should not be rejected because the texts are not 'authentic' unsimplified texts. To reject this resource is to effectively eliminate many of the essential strands of meaning-focused input and fluency development from language courses.

### REFERENCES

- Arewart, S and I S P Nation 1991. 'Fluency improvement in a second language'. *RELIC Journal*, 22, 1: 84-94

- Bauer, I. and I S P Nation 1993. 'Word families'. *International Journal of Lexicography*, 6, 4: 253-79
- Cramer, S 1975. 'Increasing reading speed in English and in the national language'. *RELIC Journal*, 6, 2: 19-23
- Day, R R and J Bamford 1998. *Extensive reading in the second language classroom*. Cambridge: Cambridge University Press
- Elley, W B 1989. 'Vocabulary acquisition from listening to stories'. *Reading Research Quarterly*, 24, 2: 174-87
- 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 and F Mangubhai 1981. *The impact of a book flood in Fiji primary schools*. Wellington: NZCER
- Ellis, R 1994. *The study of second language acquisition*. Oxford: Oxford University Press
- Joe, A 1998. 'What effects do text-based tasks promoting generation have on incidental vocabulary acquisition?' *Applied Linguistics*, 19, 3: 357-77
- Honeyfield, J 1977. 'Simplification'. *TESOL Quarterly*, 11, 4: 431-40
- Hu, M and I S P Nation 2000. 'Unknown vocabulary density and reading comprehension'. *Reading in a Foreign Language*, 13, 1: 403-30
- Krashen, S 1985. *The input hypothesis: Issues and implications*. London: Longman
- Long, M and P Robinson 1998. Focus on form: Theory, research and practice. In C Doughty and J Williams (eds), *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press
- Nation, I S P 2001. *Learning vocabulary in another language*. Cambridge: Cambridge University Press
- Nation, I S P and K Hwang 1995. 'Where would general service vocabulary stop and special purposes vocabulary begin?' *System*, 23, 1: 35-41
- Nation, P and K Wang 1999. 'Graded readers and vocabulary'. *Reading in a Foreign Language*, 12, 2: 355-80
- Newton, J 1995. 'Task-based interaction and incidental vocabulary learning: A case study'. *Second Language Research*, 11, 2: 159-77
- Nord, J R 1980. 'Developing listening fluency before speaking: An alternative paradigm'. *System*, 8, 1: 1-22
- Schmidt, R W 1992. 'Psychological mechanisms underlying second language fluency'. *Studies in Second Language Acquisition*, 14: 357-85
- Skehan, P 1996. 'A framework for the implementation of task-based instruction'. *Applied Linguistics*, 17, 1: 38-62
- Spada, N 1997. 'Form-focused instruction and second language acquisition: A review of classroom and laboratory research'. *Language Teaching*, 30, 2: 73-87
- Swain, M 1985. Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S M Gass and C G Madden (eds), *Input in second language acquisition*. Rowley, Mass: Newbury House
- West, M 1941. *Learning to read a foreign language*. London: Longman
- West, M 1953. *A general service list of English words*. London: Longman, Green & Co
- Wyldowson, H G 1976. On the authenticity of language data. In J Fanselow and R Cymes (eds), *On TESOL '76*. Washington: TESOL
- Yano, Y, M H Long and S Ross 1994. 'The effects of simplified and elaborated texts on foreign language comprehension'. *Language Learning*, 44, 2: 189-219

**Appendix**

The marked words (in bold) are not proper nouns and are not in the first 2000 words of English.

**SIMPLIFIED TEXT**

Jonathan Harker's Diary

1

Count Dracula

My story begins about seven years ago, in 1875. My name is Jonathan Harker, and I live and work in London. My job is to buy and sell houses for other people. One day a letter arrived for me from a very rich man who lived in Transylvania. He wanted to buy a house in England and he needed my help. The man was Count Dracula, and I agreed to help him. I found a house for him, and he asked me to take all the papers for it to Transylvania. I was not very pleased about this. I was planning to get married in the autumn, and I did not want to leave my beautiful Mina.

'But you must go, Jonathan,' she said. 'The Count is rich, and perhaps he will give you more work later.' So I agreed to go. I did not know then of the terrible danger which waited for me in Transylvania.

And so, on 4th May I arrived at a little town called Bistritz. Transylvania was a strange and beautiful country. There were mountains, trees and rivers everywhere. And somewhere high in the mountains was the Count's home, Castle Dracula. I had six hours to wait before the coach came to take me there, so I went into a little hotel. Inside the hotel it was warm and friendly. The people there were all laughing and talking. 'Where are you going?' they asked me. 'To Castle Dracula,' I replied.

**UNSIMPLIFIED TEXT**

**CHAPTER 1**

Jonathan Harker's Journal

3 May. Bistritz – Left Munich at 8:35 pm, on 1st May, arriving at Vienna early next morning; should have arrived at 6:46, but train was an hour late. Buda-Pesth seems a wonderful place, from the glimpse which I got of it from the train and the little I could walk through the streets. I feared to go very far from the station, as we had arrived late and would start as near the correct time as possible.

The impression I had was that we were leaving the West and entering the East; the most western of splendid bridges over the Danube, which is here of noble width and depth, took us among the traditions of Turkish

rule.

We left in pretty good time, and came after **nightfall** to Klausenburgh. Here I stopped for the night at the Hotel Royale. I had for dinner, or rather supper, a chicken done up some way with red **pepper**, which was very good but **thirsty**. (Mem. get recipe for Mina.) I asked the waiter, and he said it was called '**paprika hendl**', and that, as it was a national dish, I should be able to get it anywhere along the Carpathians.

I found my **smattering** of German very useful here; indeed, I don't know how I should be able to get on without it.

Having had some time at my **disposal** when in London, I had visited the British Museum, and made search among the books and maps in the library regarding Transylvania; it had struck me that some **foreknowledge** of the country could hardly fail to have some importance in dealing with a nobleman of that country.